


```

4251 CGATTGGCGCAGATTTGCGCAAAACCGCAGATGCGAATGGGCGTAA 4300
      ::||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1490 rAspHeaLaTyraSnValGluaSnGlnGlnInTyraSnAlaGlyLeu 1507
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
4301 ACCCGCAATCAAGGTTTCACGCTGCTCCCTCCACGCTCCGCCCAAG 4350
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1507 yLeuLysTyRHisasnValLysLeuSerLeuIleGlyGlyLeuThrLys 1523
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
4351 GGGCCGCATTTGAAGCGCAGCACAGCGGCGCATCAATTGGCTAC 4398
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
1524 AlaLysGlnAlaGlnLysGlnLysThrAlaGlnLysLeuSerPhe 1539
      ::| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|

seq_name: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:PCT-US95-10661A-3

seq_documentation_block:
: Sequence 3, Application PC/TUS9510661A
: GENERAL INFORMATION:
: APPLICANT: Washington University, et al.
: TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Flehr, Hohnach, Test, Albrighton & Herbert
: STREET: 4 Embarcadero Center, Suite 3400
: CITY: San Francisco
: STATE: California
: COUNTRY: United States
: ZIP: 94111-4187
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/10661A
: FILING DATE: 16-AUG-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/296,791
: FILING DATE: 25-AUG-1994
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Treccarlin, Richard F.
: REGISTRATION NUMBER: 31,801
: REFERENCE/DOCKET NUMBER: FP-59941/RFT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1541 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: PCT-US95-10661A-3

alignment_scores:
      Quality: 1614.00      Length: 1716
      Ratio: 1.732          Gaps: 48
Percent Similarity: 54.312 Percent Identity: 26.865

alignment_block:
US-09-303-518D-653 x PCT-US95-10661A-3 ..
Align seg 1/1 to: PCT-US95-10661A-3 from: 1 to: 1541

64 CGATTTCGCGCGGCTTACTAGCATATGCGTGTGCTTGGCATTCGCC 113
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
5 LysHeuLysLeuAsnPhelLeaLeuThrValAlaTyraLeuThrP 21
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
114 CCAAGCCCGGCGGAGCACACTTATTTGGCATCAATACCAATATCATC 163
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

21 oTyRThrcLualAlaLeuValArgAspAspValAspTyGlnIlePhea 38
164 GCGACTTTGCGCAAAATTAAGCAAGTTTGCGAGCGGGGCAAGATATT 213
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
38 rGAspPhelAlaGlnLysnGlyLysPheSerValGlyAlaThrAsnVal 54
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
214 GAGGTTTACAAACAAAAAGGAGAGTGTGGCAATTCAGTACGAAGC 263
   |||::|::|::|::|::|::|::|::|::|::|::|::|
55 LeuValLysAspLysAsnLysAspLeuGlyThrAlaLeuProAsnGly 71
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
264 C...CCGATGATGATTTTCTGTGATG...CGTAACGGCGGTGGCG 307
   :|||::|::|::|::|::|::|::|::|::|::|::|::|
71 yLlePheMetLleAspPheSerValAlaAspValAspLysArgIleAla 88
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
308 CATTCGGCGGCGATCAATATATGTCAGCGTGCACAT...AACGGCGC 354
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
88 hrLeuLleAsnProGlnTyrValValGlyAlaLysHisValSerAsnGly 104
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
355 TATAACAATGTTGATTTGGT...GCGGAGGGAAGCA 389
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
390 TCCGATCACCACCGC...TTTCTTACCAATTTG 421
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
121 nAlaLysAlaHisArgAspValSerSerGluGluAsnArgTyPheSerV 138
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
422 TGAAGAATAATTAATTAAGCAAGGACTAAGCGC... 456
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
138 aLgLuLysAsnGluTyRProThrLysLeuAsnGlyLysThrValThrThr 154
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
457 .....CATCCTATGGCGGCGATTAATCATATGCGGCTTGGCA 494
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
155 GluAspGlnThrGlnLysArgArgGluAspTyTyRMeProArgLeuAs 171
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
495 CAATTTGTGCAGATGCAGAACCTGTGAG...ATGACCAAGTTATATG 541
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
171 pLysPheValThrGlnValAlaProIleGluAlaSerThrAlaSerSerA 188
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
542 ATGGGTGCAATATACGCTGATTAATTAATACCTGATGCTGTGCAATC 591
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
188 sPAlaGlyThrTyraSnAspGlnAsnLysTyRProAlaPheValArgLeu 204
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
592 GGAGCAGGCGACACATATTTGGCGGCTGATGAAGACGA... 630
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
205 GlySerGlySerLlnPheIleTyRlysGlyAspAsnTySerLeuI 221
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
631 .CCCAATTAACCGCAAGTTCATATCAT...ATTGCAAGCG 667
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
221 eLeuAsnAsnHisGlnValGlyGlyAsnAsnLeuLysLeuValGlyAsp 238
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
668 CATATTCCTGGCGTCGCGGTGCGCAATACCTTGGCACAATAATGATCAG 717
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
238 lArTyThrTyGlyLeaAlaGlyThrProTyLysValAsnHisGluAsn 254
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
718 GGTGCGACAGCACTTAGTAGCGCAAAATAATTAACATAGC...CCATA 764
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGlnHisSerAspPro 271
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
765 TGSTTTTACCAACAGCA...GGCTCATTTGGCGACA 799
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
271 sGlyIleLeuSerGlnAspProLeuThrAsnTyAlaValLeuGlyAsp 288
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
800 GTGGCTCACCAGTATTATCTATGATGCCCAAAAGCAAAAGTGTTAAT 849
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
288 eGlySerProLeuPheValTyraSpArgGlnLysGlyLysTrpPhePhe 304
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
850 AATGGGGATTTGCAACAGCAACCCATATATAGCAAAAGCAATGGCT 899
   |||||::|::|::|::|::|::|::|::|::|::|::|::|
305 LeuGlySerTyraSpPheThrAlaGlyTyraSnLysLysSer... 318
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
900 CCAGCTAGTTGCTAAAGCTTG...TTCTATGATGAATCTTTGCTGAG 946
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
319 .....TrpGlnGluTrpAsnIleTyRlysSerGlnPheThrLysA 332
   ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

947 ATACCATTCAGTATTCTACGAACACATCAATATGGAATACTTTT 996
    |||
332 spValleu..... 334
997 AAGACAAATATATGCGAGGAAATATGATCCAAATACACTA 1046
    |||
335 .....AsnlysspserrAlaGlySerLeuIleGlySerlyThrSply 349
1047 TTCTACCTTATAGATTAAACACGACGCTT.....CAAT 1084
    |||
349 rSerThrSerSerAsnGlySerThrSerThrIleThrGlyGlyLys 366
1085 TGTTATGTTTCTTTCGAGACACACAGACAGACAGACCTGTTATCATGCT 1134
    |||
366 erLeuAsnValAspLeuAlaSpGlyLysAspLysPro..... 378
1135 GCAGGTGGGTCACAGTATTCGACCACGACTGAATATGAGAAATAT 1184
    |||
379 .....AsnHisGlyLysSerVa 384
1185 TTCTTATTATGACAAAGAAAGTGAATTGATTAATACCAACACTCA 1234
    |||
384 lThrPhe.....GluGlySerGlyThrLeuThrLeuAsnAsnIleA 399
1235 ACCAAGGCGGCGGCTTGTATTTGAGGATATTTACGATC...TCG 1281
    |||
399 spGlnGlyAlaGlyLeuPheGluGlyAspLysPyrGluValLysGly 415
1282 CCTAAAAACAAGAACGCTGGCAAGGCGGCGCTTCATATCACTGATG 1331
    |||
416 ThrSerAspAsnThrThrPlyGlyAlaGlyValSerValAlaGluG 432
1332 CAGTACCTTACTGTGAAGTAACGGCGTGGCAACGACCGCTGTCGA 1381
    |||
432 yLysThrValThrPlyValHisAsnProGlnTyThrSparGlyLeuAla 449
1382 AAATCGCAAGGACGCTGCTGCTTCAGCCAAAGGGGAAACCAAGC 1431
    |||
449 ysIleGlyLysGlyThrLeuIleValGluGlyThrGlyAspAsnLysGly 465
1432 TCGGTACCGTGGCGGAGGTAAAGTCTCTTAATCAGCAGCGGAGCA 1481
    |||
466 SerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAsnG 482
1482 TCAAGCAAAAAACAAGCCTTATGGAATCGCTGCTCAGCGCAGG 1531
    |||
482 ySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyAsn 498
1532 GGACGTCACACTGATGCGCATATATCACTCAACCCGACAACTCTAT 1581
    |||
498 erThrLeuValLeuAsnAspLysGlnValAspProAsnSerIleTy 514
1582 TTTCGCTTTCGCGGAGCTTTGATTTGAAGGGCATTCGCTTGGTT 1631
    |||
515 PheGlyPheAspGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPh 531
1632 CCACCGCATTCAAATATACGATGAAGGGCGATGTTGTCACCAACATC 1681
    |||
531 easPheIleAsnIleAsnIleAspAspGlyAlaArgLeuValAsnHisAsnM 548
1682 AAGACAAAGATCCACGCTTACCATTAACGCAATTAAGATATTACT... 1728
    |||
548 erThrAsnAlaSerAsnIleThrIleThrGlyGluSerLeuIleThrAsp 564
1728 ..... 1728
565 ProAsnThrIleThrProTyThrAsnIleAspAlaProAspGluAspAsnPr 581
1728 ..... 1728
581 cTyThrAlaPheArgArgIleLysAspGlyGlyLeuLeuThrLeuAsnLeug 598

```

```

1728 ..... 1728
598 lAsnTyThrTyThrAlaLeuArgLysGlyAlaSerThrArgSerGlu 614
1729 .....ACACCGGCATATAC..... 1743
615 leuProLysAsnSerGlyGluSerAsnGluAsnThrPleuTyMetGlyLys 631
1744 .....AACACTTGATAGCANA 1762
631 sThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnGluA 648
1763 AAGAAATGCTTACACGCTTGGTTGGCGAAGATGCAACCAACG 1812
    |||
648 rgMetAsnGlyPheAsnGlyTyThrPheGlyGluGluGlyLysAsn 663
1813 AACGGCGGCTCAATCTGAATTACCAACCGGAAGAGGGATGCACTT 1862
    |||
664 AsnGlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnArgPh 680
1863 ACTGCTTCGCGGACAAATTTAAACGGCAATATCAACCAACAAACG 1912
    |||
680 erLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGluLysG 697
1913 GCAACTGTTTTCAGCGGACAGACCGACCGCCTACATCATTTA 1962
    |||
697 lYThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAla 713
1963 GGAAGCGGCTGTCAAATGGAAGT.....ATCCCAAGAGGAAT 2006
    |||
714 GlyIleSerSerThrLysLysAspProHisPheAlaGluAsnAsnGluA 730
2007 CGTGTGGCAACAGATTTGATGACCGCATTTAAACGGCAAACTTCC 2056
    |||
730 lValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrMetA 747
2057 ATATTAGGCGGACAGCGGTGTTTC...CGCAATGTCCTCAAGTG 2103
    |||
747 snValThrGlyAsnAlaSerLeuTySerGlyArgAsnValAlaAsnIle 763
2104 GAAGCGATTTGATTTAGCAATCAATCAGCCCAAGCATTTTGGTGTCC 2153
    |||
764 ThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTy 779
2154 ACCGATCAAAAGCCACACATCTGTACAGTTTCGGATCGGAGGCTGA 2203
    |||
780 ....LysThrGlyAspThrValCysValArgSerAspTyThrGlyTy 795
2204 CAGTTGTACCGAAACCAATTCACGACAGATTAAGTTCCTTCATTG 2253
    |||
795 alThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPhe 810
2254 AGCAAGACCGACATCAAGAGCAATGTACGCTTGGCATCGCTCATTT 2303
    |||
811 AsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerAlaAsnPh 827
2304 AAATTCACAGACTTCGCCACATCAACGCGCATTTAGGACGGGAG 2353
    |||
827 eValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgGlyA 843
2354 ACACGCACTATACGTTACGCGCAACGCCACCCMAAGCGCAACTCAGC 2403
    |||
843 snSerGlnValArgLeuThrGluAsn..... 851
2404 CTCGTGGCAATGCCAAGCAACATTTAATCAAGCCATTAACGGCAA 2453
    |||
851 ..... 851
2454 CACATCGGCTTGGACATGCTTCATTTAATCAAGCAACACGCGGTAC 2503
    |||
851 ..... 851
2504 AAAAGCGACTGTGACGCTTTCGACACACGCTAAGCAACGTAAGCAT 2553

```

851 851
2554 TCGGCACTCAAGCAATGTCTCCCTAGCCGATAGGCAATTCATTT 2603
851 851
2604 TGAACACGCGCTTACCGGAAATACAGCGCGCAAGATACGCAT 2653
851 851
2654 TACACTTAAAGACAGCAATGAGCGCTCCGCGGACAGCAATTAGGC 2703
852 SerHisItrpHisIleuThrGlyAsnSerAspValHis 863
2704 AATTAAACCTTGACAAAGCCACCATTTACCTCAATTCGCGCTATGACA 2753
864 GlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaAspAsnSe 880
2754 CGATGCGCGAGCGCGCAACCGCAAGTGGCGAGATGCCCGCGCGCC 2803
880 rAsn..... 881
2804 GTTCGCGCGCTCCCTATATCCGTTACGCCCACTTCGCGAGATCC 2853
882 AsnValThr..... 884
2854 CGTTTCAACAGCTGACGGTAAACGGCAATTGACGGTCAGGAACATT 2903
885 LysIrsAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPh 900
2904 CCGGTTATGTCGGAACCTTCGCTACCGCAGCGCAATTCAGCGTG 2953
900 eTyTyIleuThrAspLeuSerAsnLysGlnGlyAspLysValValAlaI 917
2954 CGGAAGTTCCGAAGCACTTACCTTGCTGCTCAACATACCGGCAAC 3003
917 hIrsSerAlaThrGlyAsnPherThrLeuGlnValAlaAspLysThrGly 933
3004 GAACCCGTAAGTCTCGAGCAATTCAGCGTGTAGGAGAAACACACAC 3053
934 GluPro...AsnHisAsnGlnLeuThrLeuPheAspAla.....SerLy 947
3054 ACCGCTGCGAATAATCTTATTTACCTGCAAAACGAAACAGCATGATG 3103
947 SALAGlnArAspHisLeuAsnValSerLeuValGlyAsnThrValAspL 964
3104 CCGCGCATGCGCTTATCAGCTTATCCGCAAGACGCGGATCCGCGCTG 3153
964 euGlyAlaTrpLysTyIrsLeuArGAsnValAsnGlyArGtyrAspLeu 980
3154 CATATACG..... 3162
981 TyrAsnProGluValGluLysArGAsnGlnThrValAspThrAsnIl 997
3162 3162
997 eThThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnA 1014
3162 3162
1014 snGlnGluIleAlaArgValAspGluAlaProValProProAlaPro 1030
3162 3162
1031 AlaThrProSerGluThrThrGluThrValAlaGlnAsnSerLysGlnI 1047
3162 3162
1047 uSerLysThrValGluLysAsnGlnGlnAspAlaThrGluThrThrLag 1064
3163GTCAAGAACAAGAG 3177
||||| :::

1064 lAsnArgGluValAlaLysGlnAlaLysSerAsnValLysAlaAsnThr 1080
3178 CTTCGCAAACTCGGACAGCGGGA...GAACAGAGCGCGCTTGCAC 3224
1081 GlnThrAsnGluValAlaGlnSerGlySerGluThrLysGluThrGlnTh 1097
3225 GGCAAAACAGGCACAACTTCCGCCCAACACAGGCGGAAAAAGACAACG 3274
1097 rThrGluThrLysGluThrAlaThrValGluLysGluLysAlaLysV 1114
3275 CGCAAAAGC.....CTTGACCGCGCTGATTGCG 3300
1114 alGluThrLysThrGlnGlnValProLysValThrSerGlnValSer 1130
3301 GCGGCGGCAATGCCAGGAAAGCAGAAAGTGTTCGCGCAACGCGCGCG 3350
1131 ProLysGlnGlnLysSerGluThrValGlnProGlnAlaGluProAlaAr 1147
3351 GCAGCGAGCGCGGAAAAATGCCGCTTATGCGAGCGGAGAGAAAGAAA 3400
1147 g.....GluAsnAspProThrValAsnIleLysGluProGlns 1160
3401 AACGGTCAGCGCGGATTAAGACACCGCCTTG...GCGAAACAGCGGAA 3447
1160 eGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluThrSer 1176
3448 GCG..... 3450
1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193
3451GAAACCGCGCGCTTACACCGCCTTCC 3478
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209
3479 CC.....CGCGCGCGCGC 3492
1209 rThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArgArg 1225
3493 GCGCGCGCGGATTGCGG...CAACCGCAGCGCGCAACCGCAACCCACAC 3539
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242
3540 GCAGCGC.....GACCTGATCAGCGCTTATGCCAATA 3571
1242 nAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThrAsnA 1259
3572 GCGGTTGAGTGAATTTCCGCC.....ACGCTCAACAGC 3606
1259 lValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275
3607 GTTTTCGCGTACAGGAGCAATTCGACGCGGTG...TTTCCGGAAGACCG 3653
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGlnMetAsnAsnGlnI 1292
3654 CCGCAACGCGCTTGTGACAGCGCATCCGGGACACCAACACTACCGTT 3703
1292 yGlnTyrsAsnValTrpValSerAsnThrSerMetAsnLysAsnTyrsS 1309
3704 CCGCAATTCGCGCGCTTACCGCAACAAACGACCTGCGCAATTCGCT 3753
1309 eSerGlnTyrsArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325
3754 ATGCAAGAAACCTCGCAGCGGG...GCGTCCGCGATCCCTTTTTCGCA 3800
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1342
3801 CAACGCGACCGGAAACACCTTCGACAGCGCATTCGCAATCGGACGCGC 3850
1342 rValArgAsnSerAsnAsnPhAspLysAlaThrSerLysAsnThrLeuA 1359
3851 TTGCCCAGCGTCCGCTTTCGGGCAATACGCAATCGCAGGTTCCGACATC 3900
1359 lAcIlnValAsnPhetyrSerLysTyTyTrpAlaAspAsnHisTrpTyLeu 1375

```

: TELEPHONE: (415) 781-1989
: TELEFAX: (415) 398-3249
: TELEX: 910 277299
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1545 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: US-08-296-791-4

alignment_scores:
    Quality: 1587.00      Length: 1723
    Ratio: 1.710          Gaps: 47
    Percent Similarity: 53.860    Percent Identity: 26.291

alignment_block:
US-09-303-518D-653 x US-08-296-791-4 ..

Align seg 1/1 to: US-08-296-791-4 from: 1 to: 1545

64 CGTTTCTGCGCCGGCTACTTATGACCATATGCTGCTTGCGCATTTGCG 11
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5 LysPheIysLeuAsnPhelIeAlaLeuThrValAlaIryAlaLeuThrPr 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
114 CCAAGCCCGGGGGGACACACTTATTTGCGATCACTACCAATACATATC 165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
21 CTYrThrGluAlaIaLeuValAlaGspAspValAlaSPyrGlnIlePheA 38
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
164 GCGACTTTGGCGAAATATAAAGCGAAGTTTGCGATCGGGGCGGAAGATAT 21
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
38 rGspPheAlaIaGluAsnIysGlySPheSerValAlaIryAlaThrAsnVal 54
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
214 GAGGTTTACACAACAAAAAGGGGAGTGTGTCGCGAATCGATGACGAAAGC 263
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
55 GluValAlaGspIysAsnAsnArgProLeuGlyAsnValLeuProAsnG 71
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
264 C...CCGATGATTATTTTCTGTGATATGC..CGTACGGCGTGGCGG 307
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
71 yIlePrrMetIleAspPheSerValAlaAspValaSPyrAlaIryThrV 88
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
308 CATTTGGCGGGCGGATCAATATTTGTGACGGGCGACAT...AACGGGCG 354
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
88 hrIeuValaSPyrGlnIryValAlaGlyAlaIryLysIValSerAsnGly 104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
355 TATAACAGATGTTGATTTTGGT.....GCGAGGAGAACAA 389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
390 TCCCGATCAGACAGCC.....TTTTCTTACCAATTTG 422
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 naIaLysAlaIaHisIrrGspValSerSerGluGluAsnArgIryThrV 138
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
422 TGAAGAACAATATTAATTAAGAAGACAGGACTTAAGCGC.....456
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
138 alGluIysAsnGluIryProThrIryLysLeuAsnGlyIryAlaValaIThr 154
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
457 .....CATCTTATGCGCGGATTAATCAATATGCGCGTTTGCA 494
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
155 GluAspGlnAlaGlnIysArgArgGluAspIryTyrMetProArgLeuAs 171
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
495 CAATTTGTCAAGATGACAGACAGACTGTTGAG..ATGACAGATTATATG 541
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
171 pIysPheValIrrGluValAlaProIleGluAlaSerThrAspSerSerT 188
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
542 ATGGGTTGAATATCGCTGATTAATTAATTAATACCTGATCGTTGCGATC 591
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
188 hrIaGluIryThrIryAsnAsnIysAspIryTyrProIryPheValArgLeu 204
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
592 GGAGCAGCGACACAAT.....TG 611
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
205 GlySerGlyThrGlnPheIleIryGluAsnGlyThrArgTyrGluLeuThr 221

```



```
777 alyValHisIleGlyTyr.....LysAlaGlyAspThrValCysVala 792
2183 GTTGGAGCTGGAGCGGTGTGACAGTGTGTCCGCAAAAAACCATTAACGGAC 2232
      |||||:::|||||:::|||||:::
792 rGSeAspTyrThrGlyTyrValThrCysThrThrAspLysLeuSer... 807
2233 GATTAAGGATGCTGCTATTGAGCAAGACGACATCAGAGGCAATGTCAAG 2282
      |||||:::|||||:::|||||:::
808 AspLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAsnValAs 824
* 2283 CATTGCGGATCAGCGCTCATTTAAATCTCACAGACTTGCACACTCAACG 2332
      |||||:::|||||:::|||||:::
824 nLeuThrGluSerAlaAsnPheValLeu...GlyLysAlaAsnLeuPheG 840
2333 GCATCTTAGTGCAGCGGACACACGCACTATAGGTTACGGCGCAAGGCC 2382
      |||||:::|||||:::|||||:::
840 LyrThrIleGlnSerArgGlyAsnSerGlnValArgLeuThrGluAsn... 855
2383 ACCCAAAACGGCAACCTCAGCGCTGTGGGCAATGCCAAGCAACATTAA 2432
855 .....
2433 TCAAGCCACATTAAACGGCAACACATCGGCTCGACAAATGCTTCAATT 2482
855 .....
2483 ATCTAAGCAACACGCGCTACAAAACGGCACTGACGCTTCGGACAC 2532
855 .....
2533 GCTAAGCAACAGTAAGCCATTCCGCACTCAACGGCAATGTCTCCCTAGC 2582
855 .....
2583 CGATTAGGCAATTCATTCTTTGAAACAGCGCTTACCGGAAATCA 2632
855 .....
2633 GCGGCGGCAAGATACGCGATTACACTTAAAGACAGCAATGAGCGGTG 2682
856 .....SerHisTrpHisLeu 860
2683 CCGTGGGCGACGGAATTAGGCAATTAAACCTTGACAAACGCGCACTTAC 2732
      |||||:::|||||:::|||||:::
861 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlyHisIleH 877
2733 ACTCAATTCGCGCTATGACACAGATGGCGGCGGCAACCGGCAGTG 2782
      |||||:::|||||:::
877 sLeuAsnSerAlaAspAsnSerAsn..... 885
2783 CGGCAATGCGCGCGCGCGCTTCGCGCTCCCTATATCCGTTACG 2832
886 .....AsnValThr 888
2833 CCGCAACTTGGCGAGAAATCCGTTTACACGCTGACGCTAAACGGCA 2882
      |||||:::|||||:::|||||:::
889 .....LysTyrAsnThrLeuThrValAsn...Se 897
2883 ATTGAAGCGTACGAGCAACATCCGTTATATGCGAATCTTCGGCTAC 2932
      |||||:::|||||:::|||||:::
897 rLeuSerGlyAsnGlySerPheTyrTyrLeuThrAspLeuSerAsnLysG 914
2933 GCAGCGGCAAAATTGAAGCTGGCGGAAAGTCCGAAGGCACTTACACTTG 2982
      |||||:::|||||:::|||||:::
914 InGlyAspLysValValValThrLysSerAlaThrGlyAsnPheThrLeu 930
2983 GCTGTCAACATATCCGCGACGACCGCTAAGTCTCGACAAATGACGCT 3032
      |||||:::|||||:::|||||:::
931 GlnValAlaAspLysThrGlyLysPro...AsnHisAsnGluLeuThrLe 946
3033 AGTGAAGGAAAGACAAACACACCGCTGTCGAAATATCTAATTTACACC 3082
      |||||:::|||||:::|||||:::
946 uPheAspAla.....SerLysAlaGlnArgAspHisLeuAsnValSerL 961
3083 TGCAAAACGAAACAGCTGCATGCCGCGCATGCGCTTATCAGCTTATCCGC 3132
      |||||:::|||||:::|||||:::
961 euValGlyAsnThrValAlaAspLeuGlyAlaTrpLysTyrLysLeuArgS 977
3133 AAGACGCGGAGTCCGCTGCATTAATCCG..... 3162
      |||||:::|||||:::|||||:::
978 ValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsnG 994
3162 ..... 3162
994 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1011
3162 ..... 3162
1011 alProSerValProSerAsnAsnGluGluIleAlaArgValAspGluAla 1027
3162 ..... 3162
1028 ProValProProProAlaProAlaThrProSerGluThrThrGluThrVa 1044
3162 ..... 3162
1044 lAlaGluAsnSerLysGlnGluSerLysThrValGluLysAsnGluGlnA 1061
3162 ..... 3162
1061 sPalaThrGluThrThrAlaGlnAsnArgGluValAlaLysGluAlaLys 1077
3163 .....GTCAAAGACAAAGACGCTTCCGCAACACTCGGCAAGCGGGA... 3204
      |||||:::|||||:::|||||:::
1078 SerAsnValLysAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySe 1094
3205 .GAACAGAGCGCGCTTACGCGCAAAACAGCACACTTGCAGCGCAAC 3253
      |||||:::|||||:::|||||:::
1094 rGluThrLysGluThrGlnThrThrThrLysGluThrAlaThrValG 1111
3254 AACAGCGGAAAAAGACAAACGCGCAAGC..... 3282
      |||||:::|||||:::|||||:::
1111 LuLysGluGluLysAlaLysValGluThrGluLysThrGlnGluValPro 1127
3283 ...CTTGACGCGCTGATTCGCGCGCGCGCATGCCACGCAAAAGCGCA 3329
      |||||:::|||||:::|||||:::
1128 LysValThrSerGlnValSerProLysGlnGluGlnSerGluThrValG 1144
3330 AAGTGTTCGCAACCGCGCGCGCGAGCGGCAAGGAAATGCCGCACTTA 3379
      |||||:::|||||:::|||||:::
1144 nProGlnAlaGluProAlaArg.....GluAsnAspProThrV 1157
3380 TGCAGCGGAGAGAGAGAAAAAGGCTGCAGCGGATTAAGACACCGCC 3429
      |||||:::|||||:::|||||:::
1157 alAsnIleLysGluProGlnSerGlnThrAsnThrAlaAspThrGlu 1173
3430 TTG...GCCAAACAGCGCGAAGCG..... 3450
1174 GlnProAlaLysGluThrSerSerAsnValGluGlnProValThrGluSe 1190
3451 .....GAAACCC 3457
1190 rThrThrValAsnThrGlyAsnSerValValGluAsnProGluAsnThrT 1207
3458 GCGCGGCTACACCGCTTCCG..... 3480
1207 hrProAlaThrThrGln...ProThrValAsnSerGluSerSerAsnLys 1222
3481 .....CGCGCCGCGCGCGCGCGGATTTGCG...CAACCGCA 3518
1223 ProLysAsnThrGlnArgGlySerValArgSerValProHisAsnValG 1239
3519 GCCCAACCGCAACCCCAACCGCAGCGC.....GACC 3550
      |||||:::|||||:::|||||:::
1239 uProAlaThrThrSerSerAsnAspArgSerThrValAlaLeuCysAspL 1256
```



```

308 CATTCGGCGGCGATCATATATATGAGCGTGGACAT...ACGGCGGC 354
   ::::::::::: |||:::|||||::: ||| ::::::::::: |||
88 hrlleuValasnProglntYrValValGlyValLysHisValSerAsnGly 104
355 TATAACAATGTTGATTTTGGT.....GGGAGGGAGCA 389
   ::::::::::: |||:::
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnGlyAs 121
390 TCCCATGACGACCCG.....TTTCTTACCAAAATG 421
   |||::: |||
121 hAlaLysAlaHisArgAspValSerSerGluLysAsnArgTrpTrpThr 138
422 TGAAGAAGAAATATATTAAGCAGGAGCTAAGCGC..... 456
   |||::: |||::: |||
138 alGluLysAsnGluTrpProThrTrpLysLeuAsnGlyLysAlaValThrThr 154
457 .....CATCCTTATGCGCGGATTAATCATATGCGCGTTTGA 494
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||
155 GluAspGlnAlaGlnLysArgArgGluAspTrpTrpMetProArgLeuAs 171
495 CAAATTTGTCACAGATGCAAGACCTGTGAG...ATGACGATATATAG 541
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
171 pLysPheValThrGluValAlaProLeuAlaSerThrAspSerSert 188
542 ATGGGTGAAATACGCTGATTAATTAATACCTGATCGTGTGCAATC 591
   ||| :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
188 hAlaGlyThrTrpAsnAsnLysAspLysTrpProTrpPheValArgLeu 204
592 GGAGCAGCGACAAATAT.....TG.611
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
205 GlySerGlyThrGlnPheIleTrpGluAsnGlyThrArgTrpGluLeuTr 221
612 GCGGTCTGATGAGCAACCAATACCGGAAGTTCATATCAT... 657
   | :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
221 pLeuGlyLysGlnGlnLysSerAsp...AlaGlyGlyTrpAsnLeuL 237
658 .....ATTGCAAGCGCATATCTTGTCGTGCGTGGCGCAATACCTTGCA 702
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
237 yLsLeuValGlyAsnAlaTrpThrTrpGlyLysIleAlaGlyThrProTrpGlu 253
703 CAAATGATCAGTGTGGCGACAGTCAACTTAGTAC..... 741
   ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: |||
254 ValAsnHisGluAsnAspGlyLeuIleGlyPheGlyAsnSerAsnAsnGly 270
742 .....GAAAAATTAACATAGCCCATATGTTTATAC 775
   ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: |||
270 uTrpIleAsnProLysGluIleLeuSerLysLysPro.....LeuT 284
776 CAACAGAGGCTCATTTGGCGACAGTGGCTCACAATGTTATCTATGAT 825
   :::: :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
284 hAsnTrpAlaValLeuGlnLysSerGlySerProLeuPheValTrpAsp 300
826 GCCCAAAAGCAAAAGTGTATATTAATGGGATTCGCAAAACAGCAACCC 875
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
301 ArgGluLysGlyLysTrpLeuPheLeuGlySerTrpAspTrpTrpAlaGly 317
876 CTATATAGCAAAAACCAATGCTTCACGCTAGTTCGTAAGATGTTGTTCT 925
   ||| :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
317 yTrp.....AsnLysLysSerTrpGlnG 325
926 ATGATGAATCTTTGCTGAGATACCATTCAGTATTTCTAGCAACACAT 975
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
325 LuTrpAsnIleTrpLysProGluPheAlaGlnLysIleTrpGlnGlnTrp 341
976 CAAATGGGAATACTTTTAAACACATAATATGCGCAGAGAAAT 1025
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
342 .....SerAlaGlySerLe 346
1026 CGATGCCAAACATAAACAATACTTCTACTTATAGATTAAACACCA 1075
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
346 uIleGlySerLysThrAspTrpSerSerAsnGlyLysThrSert 363
1076 CCGTT.....CAATGTTTAATGTTTCTTATCCGAGACACCA 1113
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
363 hrlleuThrGlyGlyGluLysSerLeuAsnValAsnLeuAlaAspGlyLys 379
1114 AGAAGACCTGTTTATCATGCTGCAGGTGGGTCACAGTTATTCAGCCAG 1163
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
380 AspLysPro..... 382
1164 ACTGAATTAATGAGAAATATTTCTTTATGCAAAAGCAAAAGCTGAT 1213
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
383 ....AsnHisGlyLysSerValThrPhe.....GluGlySerGlyThrL 396
1214 TGATACTTACACACATCAACCAAGCGCGCGGCTGTGATTTGAG 1263
   ||| :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
396 euthrLeuAsnAsnAlaAsnIleAspGlnGlyAlaGlyLysLeuPheGln 412
1264 GGTAAATTTTACGGTC...TCGCTAAAAACACAGAAAGCTGCGACGGCC 1310
   |||::: ||| :::: |||::: |||::: |||::: |||::: |||::: |||
413 GlyAspTrpGluValLysGlyThrSerAspAsnThrThrTrpLysGlyAl 429
1311 GGGCGTTTCATATCATGATGACGATACCGTTACTTGAAGATTAACGGCG 1360
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
429 alGlyValSerValAlaGlnGlyLysThrValThrTrpLysValHisAsn 446
1361 TGGCAAAACGACCGCTGTCCAAATCGCAAGCAGCGCTGCTTCAA 1410
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
446 roGlnTrpAspArgLeuAlaLysIleGlyLysGlyThrLeuIleValGlu 462
1411 GCCAAAGGGGAAACCAAGGCTGGTCAGCGTGGCGGAGCGTAAGTCAT 1460
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
463 GlyThrGlyAspAsnLysGlySerLeuLysValGlyAspGlyThrValAl 479
1461 CTATGATCAGCAGCGAGCAGATCAAGCAAAACCAAGCCTTATGTA 1510
   ||| :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
479 eleuLysGlnGlnThrAsnGlySerGly...GlnHisAlaPheAlaSerY 495
1511 TCGGCTTGGTCAAGCGGAGGGGAGCGTGCACACTGATCCGATAATCG 1560
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
495 alGlyIleValSerGlyArgSerThrLeuValLeuAsnAspLysGln 511
1561 TTCACCCGCAAAACCTATTTGCGCTTGGCGGCGGAGCTTGAATTT 1610
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
512 ValAspProAsnSerIleTrpPheGlyPheArgGlyValArgLeuAspLe 528
1611 GAACGGGCAATTCGTTTGTTCACCGCATTCAAATACCGATGAGAGGG 1660
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
528 uAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspGlnGlyA 545
1661 CGATGATTTGCAACCAATCAAGACAAAGAATCCACCGTTACATTA 1710
   ||| :::: |||::: |||::: |||::: |||::: |||::: |||::: |||
545 IaArgLeuValAsnHisSerThrSerLysHisSerThrValThrIleThr 561
1711 GGCATTAAGAATTTACTACACCGGCAT..... 1740
   ::::::::::: |||::: |||::: |||::: |||::: |||::: |||::: |||
562 GlyAspAsnLeuIleThrAspProAsnAsnValSerIleTrpTrpValY 578
1740 ..... 1740
578 sProLeuGluAspAsnProTrpAlaIleArgGlnIleLysTrpGlyT 595
1740 ..... 1740
595 yGlnLeuTrpPheAsnGlnGluAsnArgThrTrpTrpAlaLeuLysLys 611
1740 ..... 1740
612 AspAlaSerIleArgSerGluPheProGlnAsnArgGlyGluSerAsnAs 628
1741 .....A 1741
628 uSerTrpLeuTrpMetGlyThrGluLysAlaAspAlaGlnLysAsnAla 645
1742 ACAACAACCTTGATAGCAAAAGAAATTGCCCTACACAAGCTTGTTGGC 1791
   |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||

```



```
645 etasnHisIleasnAsnGluArgMetAsnGlyPheasnGlyTyrPheGly 661
1792 GAGAAAGATCGCAACAAACGAGGGCGCTCAATCTGTAATTAACAAC 1841
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
662 GluGluGluGly...LysAsnAsnGlyAsnLeuAsnValThrPheLysG1 677
1842 GGAAGAACGGATCGCACTTACTGCTTCCGGCGGACAAATTTAAAG 1891
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
677 LysSerGluGlnAsnArgPheLeuThrGlyGlyThrAsnLeuAsnG 694
1892 GCATATACGCAACAAACGCAAACTGTTTTCAGCGGACAGCCGACA 1941
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
694 LysPheLeuAsnValGlnGlnGlyThrLeuPheLeuSerGlyArgProTh 710
1942 CCCGACGCTACAACTATTAGAGAGGGGTGTCAAAAATGAGAGT... 1989
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
711 ProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAspSerH1 727
1990 ....ATCCCAAGAGAGAAATCGTGTGGACAAAGATTTGATCGACGCA 2035
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
727 sPheSerGluAsnAsnGluValValValGluAspAspTrpIleAsnArgA 744
2036 CATTTAAAGCGGAAACTTCATATTCAGCGCGGACAGCGGTGTTCC 2085
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
744 snPheLysAlaThrAsnIleAsnValThrAsnAsnAlaThrLeuTyrSer 760
2086 ...CGCAATGTGCAAAAGTGAAGGAGATTCGATTTAAGCAATGACGC 2132
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
761 GluArgAsnValGluSerIleThrSerAsnIleThrAlaSerAsnAla 777
2133 CCAAGCAGTTTTCGGTGTGCGACCGCATCAACACACAACTGTACAC 2182
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
777 aLysValHisIleGlyTyr....LysAlaGlyAspThrValCysValA 792
2183 GTTCGACATGACGGCTGTACAGATTGTACGAAAAAACCATTTACCGAC 2232
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
792 rGserAspTyrThrGlyTyrValThrCysThrThrAspLysLeuSer... 807
2233 GATAAAGTATGCTTCATTCATGACAGACGACATCAGAGCAATGTCAG 2282
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
808 AspLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAsnValA 824
2283 CCTTCGCGATCAGCGTCATTAAATCTCACAGGACTTGCACACTCAAG 2332
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
824 nLeuThrGluSerAlaAsnPheValLeu...GlyLysAlaAsnLeuPheG 840
2333 GCATCTTAATGCAAGCGGACAGACGCACTATACGTTACGCCAAGCC 2382
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
840 LysThrIleGlnSerArgGlyAsnSerGlnValArgLeuThrGluAsn... 855
2383 ACCCAAAAGCGCAACCTCAGCCTCGTGGGCAATGCCAAGCAACATTAA 2432
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
855 ..... 855
2433 TCAAGGCACATTAAACGGCAACATCGGCTTCGACAATGCTTCATTAA 2482
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
855 ..... 855
2483 ATCTAAACAACAGCCGCTACAAACGCGAGTGTGACGCTTTCGACAAAC 2532
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
855 ..... 855
2533 GCTAAGCAAGCTAAGCATTCGCGACTCAACGGCAATGTCTCCTACG 2582
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
855 ..... 855
2583 CGATTAAGCAGATTCCATTTTGAAAACAGCCGCTTACCGGAAAAATCA 2632
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
855 ..... 855
2633 GCGGCGCAAGGATACGGCATTCACATTAAAGACAGCGAATGACGCTG 2682
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
856 .....SerHisTrpHisLeu 860
2683 CCGTGGCGACGAATTAAGCAATTTAAACCTTGCAACAGCCACCATAC 2732
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
861 ThrGlyAsnSerAspValHisGlnLeuAspLeuAlaAsnGlyHisIleH1 877
2733 ACTCAATTCGCGCTTTCGACAGATCGCGCAGCGCGCAACCGGCACTG 2782
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
877 sLeuAsnSerAlaAspAsnSerAsn..... 885
2783 CCGCAGATGCGCGCGCGCGCTTCGCGCGCTTCCCTTATTCCGTACG 2832
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
886 .....AsnValThr 888
2833 CCGCCACATTCGCGAGAAATCCGCTTCAACACGCTGACGTAACGGCAA 2882
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
889 .....LysTyrAsnThrLeuThrValAsn...Se 897
2883 ATTGAACGCTCAGGAGACATTCGCTTATGTCGACACTTCGCTAC 2932
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
897 rLeuSerGlyAsnGlySerPheTyrTyrLeuThrAspLeuSerAsnLysG 914
2933 GCAGCGGCAATTTGAAGCTGCGGAAAGTTCCGAAGGCACTTACACTTG 2982
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
914 InGlyAspLysValValValThrLysSerAlaThrGlyAsnPheThrLeu 930
2983 GCTGTCAACAATACCGGACAGCAACCGCTAAGTCTCGACAAATGACGT 3032
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
931 GlnValAlaAspLysThrGlyGluPro...AsnHisAsnGluLeuThrLe 946
3033 AGTGAAGAGAAAAGACAAACACACCGCTGTCGAAATCTTATTTACACC 3082
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
946 uPheAspAla....SerLysAlaGlnArgAspHisLeuAsnValSerL 961
3083 TCCAAAGACACACGTCGATGCGCGCATGCGCTTATCAGCTTATCCGC 3132
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
961 euValGlyAsnThrValAspLeuGlyAlaTrpLysLysLeuArgAsn 977
3133 AAGACGGGAGATCCGCTGCATATACG..... 3162
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
978 ValAsnGlyArgTyrAspLeuTyrAsnProGluValGluLysArgAsnG1 994
3162 ..... 3162
994 nThrValAspThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspV 1011
3162 ..... 3162
1011 aLProSerValProSerAsnAsnGluGluIleAlaArgValAspGluAla 1027
3162 ..... 3162
1028 ProValProProProAlaProAlaThrProSerGluThrThrGluThrVa 1044
3162 ..... 3162
1044 lAlaGluAsnSerLysGlnGluSerLysThrValGluLysAsnGluGlnA 1061
3162 ..... 3162
1061 sPAlaThrGluThrThrAlaGlnAsnArgGluValAlaLalysGluAlaLys 1077
3163 .....GTCAAAGAACAGAGCTTCGACAAACCTCGGCAAGGGGGA... 3204
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1078 SerAsnValLysAlaAsnThrGlnThrAsnGlnValAlaGlnSerGlySe 1094
3205 GAACAGAGCGCGCTTACAGCGCAAAACAGCAACACTTCGCGCAAAAC 3253
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1094 rGluThrLysGluThrGluThrThrGluThrLysGluThrAlaThrValG 1111
3254 AACAGCGGAAAAAGACAACGCGCAAGC..... 3282
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1111 lLysGluGluLysAlaLysValGluThrGluLysThrGlnGluValPro 1127
```

```

3283 ...CTTGACGGCTGATGTCGGCGCGGCAATGCCAAGCAAGGAGGAGA 3329
      ::::: ::::: ::::: ::::: :::::
1128 LysValThrSerGlnValSerProLysGlnGlnSerGlnThrValGln 1144
      ::::: ::::: ::::: ::::: :::::
3330 AAGTGTCCCAACCGCGCGCGGCAAGCGCGGGAATGCCGGCATTA 3379
      ::::: ::::: ::::: ::::: :::::
1144 nProGlnAlaGlnProAlaArg.....GluAsnAspProThrV 1157
      ::::: ::::: ::::: ::::: :::::
3380 TGCACGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3429
      ::::: ::::: ::::: ::::: :::::
1157 AlaSerGlnValSerGlnThrSerGlnThrValAlaSerProGln 1173
      ::::: ::::: ::::: ::::: :::::
3430 TTG...GCGAACAACGCGGAGCG..... 3450
      ::::: ::::: ::::: ::::: :::::
1174 GlnProAlaLysGlnThrSerSerAsnValGlnGlnProValThrGln 1190
      ::::: ::::: ::::: ::::: :::::
3451 .....GAAACCC 3457
      ::::: ::::: ::::: ::::: :::::
1190 rThrThrValAsnThrGlnAsnSerValValGlnAsnProGlnAsnThr 1207
      ::::: ::::: ::::: ::::: :::::
3458 GCGCGGTACACCGCGCTTCCCG..... 3480
      ::::: ::::: ::::: ::::: :::::
1207 hrProAlaThrThrGln...ProThrValAsnSerGlnSerSerAsnLys 1222
      ::::: ::::: ::::: ::::: :::::
3481 .....CGCGCGCGCGCGCGCGCGGATTTGCCG...CAACGCA 3518
      ::::: ::::: ::::: ::::: :::::
1223 ProLysAsnArgHisArgArgSerValArgSerValProHisAsnValGln 1239
      ::::: ::::: ::::: ::::: :::::
3519 GCGCCACCGCAACCCCAACCGCAACCGC.....GACC 3550
      ::::: ::::: ::::: ::::: :::::
1239 nProAlaThrThrSerSerAsnAspArgSerThrValAlaLeuGlnAspL 1256
      ::::: ::::: ::::: ::::: :::::
3551 TGATCAGCGCTTATGCCAATAGCGTTGATGATGATTTCCGCC..... 3594
      ::::: ::::: ::::: ::::: :::::
1256 euThrSerThrAsnThrAsnAlaValLeuSerAspAlaArgAlaLysAla 1272
      ::::: ::::: ::::: ::::: :::::
3595 .....ACGCTCACACAGCGTTTCCGCGTACAGAGACGATTTGACCG 3635
      ::::: ::::: ::::: ::::: :::::
1273 GlnPheValAlaLeuAsnValGlnLysAlaValSerGlnHisIleSerGln 1289
      ::::: ::::: ::::: ::::: :::::
3636 CGTG...TTTGCGGAGACGCGCGGCAACGCGGTTGGACACGCGGCATCC 3682
      ::::: ::::: ::::: ::::: :::::
1289 nLeuGlnLeuAsnAsnGlnGlnGlnLysLysLysLysLysLysLysLys 1306
      ::::: ::::: ::::: ::::: :::::
3683 GGGACACCAACAACTACTACCGTTCCGAGATTTCCGCGCTACGCCACAA 3732
      ::::: ::::: ::::: ::::: :::::
1306 erMetAsnLysAsnLysSerSerSerGlnLysArgAlaArgPheSerSerLys 1322
      ::::: ::::: ::::: ::::: :::::
3733 ACCGACCTGCGCCAAATCGGTATGCAGAAAAAAGCTGCGGACGGG...CG 3779
      ::::: ::::: ::::: ::::: :::::
1323 SerThrGlnThrGlnLeuGlnLysPheGlnThrIleSerAsnAsnValGln 1339
      ::::: ::::: ::::: ::::: :::::
3780 CGTGCGGACCTCGTTTGGCAACCGCGGAGCAACCTTCGACGACGCG 3829
      ::::: ::::: ::::: ::::: :::::
1339 nLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1356
      ::::: ::::: ::::: ::::: :::::
3830 GCATCGGCAACTCGGACGCGGTTGCGGCGGCGGTTTGGCGCAATAC 3879
      ::::: ::::: ::::: ::::: :::::
1356 IaThrSerLysAsnThrLeuAlaGlnValAsnPheLysSerLysLysLys 1372
      ::::: ::::: ::::: ::::: :::::
3880 GGCATCGGACGAGTTGCATCGGACATCGGCGGCGGCGGCGGTTTGTAGTAG 3929
      ::::: ::::: ::::: ::::: :::::
1373 AlaAspAsnHisLysLysLysLysLysLysLysLysLysLysLysLys 1389
      ::::: ::::: ::::: ::::: :::::
3930 CGGACGCTTTCAGACGCGCATCAGAGGCAAAATCCCGCGCGCGCTGCTGC 3979
      ::::: ::::: ::::: ::::: :::::
1389 nSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1406
      ::::: ::::: ::::: ::::: :::::
3980 ATTACGATTCAGGACGAGATACGCGCGGCGGCGGCGGATTCGGGATC 4029
      ::::: ::::: ::::: ::::: :::::
1406 LnpheGlnLysThrAlaGlnLysLysLysLysLysLysLysLysLysLys 1422
      ::::: ::::: ::::: ::::: :::::
4030 GAACGCGACATCGGCGCAACGCGGATTTCTCCAAAAAGCGGATTCGCG 4079

```

```

1143 ThrProIleValGlnValArgLysSerLysLysSerAsnAlaAspPheAl 1439
      ::::: ::::: ::::: ::::: :::::
4080 ATACGAAACGTCATATATGCGACCCCGGCGCTTGATTCACCGGTACC 4129
      ::::: ::::: ::::: ::::: :::::
1439 AlaAspGlnAlaArgLysLysValAsnProIleSerValLysThrAlaP 1456
      ::::: ::::: ::::: ::::: :::::
4130 GCGCGGCGCATTAAGCAGATTTATTCATCAAAACCGCGCAACATTTCC 4179
      ::::: ::::: ::::: ::::: :::::
1456 heAlaGlnValAspLysSerLysLysLysLysLysLysLysLysLys 1471
      ::::: ::::: ::::: ::::: :::::
4180 ATCAGCGCTTATTAAGCGCTGCTCTATACGATGCCGCTTCGCGCAAGT 4229
      ::::: ::::: ::::: ::::: :::::
1472 ValThrProIleLysSerAlaArgLys...AspAlaAsnGlnLysSerGln 1487
      ::::: ::::: ::::: ::::: :::::
4230 CGGACGCGGCGTCATACCGCGGCTATTTGGCGGAGATTTCCGCAAAACC 4279
      ::::: ::::: ::::: ::::: :::::
1487 LysLysLysValAsnGlnLysLysLysLysLysLysLysLysLysLys 1504
      ::::: ::::: ::::: ::::: :::::
4280 GCAATGCGGATGAGGCGGTAAACGCCCAATCAAGGTTTCAGCGGTGCC 4329
      ::::: ::::: ::::: ::::: :::::
1504 LngLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1520
      ::::: ::::: ::::: ::::: :::::
4330 CTCACGCTGCGCGCGCAAGGCGCGCAATTTGAAAGCGGACACAGCGC 4379
      ::::: ::::: ::::: ::::: :::::
1521 LeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1537
      ::::: ::::: ::::: ::::: :::::
4380 GGGCATCAATTAAGGCTAC 4398
      ::::: ::::: ::::: ::::: :::::
1537 aGlnLeuLysLysSerPhe 1543

```

seq_name: /cgn2_6/ptodata/1/1ae/6b_COMB.pep:US-08-296-791-5

seq_documentation_block:

Sequence 5, Application US/08296791

Patent No. 6243337

GENERAL INFORMATION:

APPLICANT: SL. Game III, Joseph W.

TITLE OF INVENTION: Haemophilus Adherence and Penetration

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/296-791

FILING DATE: 25-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Treacartin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 1702 amino acids

TYPE: amino acid

TOPOLOGY: unknown

US-08-296-791-5

[illegible]

1004	rthrrProAsnAsnIIeGlnAlaAspValProSerValProSerAsnAsnG	1021
3162	3162
1021	IuGluIleAlaArGValGIuThrProValProProAlaProAlaThr	1037
3162	3162
1038	ProSerGIuThrThrGIuThrValAlaGluAsnSerIysGIuSerLy	1054
3163GTCAAAGACAAGAGCTTCCGACAACCTCGCAAGCGG	3202
1054	sthValGIuIyrsAsnGIuGlnAspAlaThrGIuThrThrAlaGlnAsnG	1071
3203	GAGAA.....	3207
1071	IyGIuValAlaGIuGlnAlaIyProSerValIyAlaAsnThrGIuThr	1087
3208ACAGA	3212
1088	AsnGIuValAlaGIuSerGIySerGIuThrGIuThrGlnThrThrGI	1104
3213	GGCCGCGCTGACGGCA.....	3228
1104	uIeIyrsGIuThrAlaIyValGIuGlnSerGIuIyrsAlaIyValGIuI	1121
3228	3228
1121	ysGIuIyrsAlaIyValGIuIyrsAspGIuIeGIuIyAlaProGln	1137
3229AAACAGCGACACTTGGCGGCAAA.....	3252
1138	MetAlaSerGIuThrSerProIyGlnAlaIyProAlaProIyrsGIuVa	1154
3252	3252
1154	IserThrAspThrIyrsValGIuIuThrGlnValGlnAlaGlnProGlnT	1171
3252	3252
1171	hrcIuSerThrThrValAlaAlaGluAlaThrSerProAsnSerIys	1187
3253CAACAGCGGAAAAAGACAACGCGCAAGCGCTTGA	3287
1188	ProAlaGIuIuThrGlnProSerGIuIyrsThrAsnAlaGIuProValTh	1204
3288	CGCGCTGATTGGCGCGCGCAAT.....	3312
1204	rProValIySerIyrsAsnGIuThrGIuAsnThrThrAspIuProThrG	1221
3313GCCACGGAAAAAGCAAGCTTGCC	3339
1221	IuArGIuIysthrrAlaIyrsValGIuThrGIuIysthGlnIuProPro	1237
3340	GAACCGCGCGGACGAGCAGC.....	3360
1238	GlnValAlaSerGIuAlaSerProIyrsGlnGIuSerGIuThrValGI	1254
3361GGGGAATAATGCC...GGCATTTATGACGCGG	3388
1254	nProGlnAlaValLeuGIuSerGIuAsnValProThrValAsnAsnAlaG	1271
3389	AGCAAGACAAAAACGGGTGACGGCGATAAAGAC.....	3423
1271	IuGIuValGIuAlaGIuLeuGIuThrGlnThrSerAlaThrValSerThr	1287
3424ACCGCCTT	3431
1288	IyGIuIuProAlaProGIuAsnSerIleAsnThrGIySerAlaThrAlaI	1304
3432	GGGGAACAGCGCGAAGCGGAACCCGGCGGCTACC.....	3468

1304	eThrGluThrAlaGluLysSerAspLysProGlnThrGluThrAlaAsp	13321
3468	3468
1321	eThrGluLysPalaSerGlnHisLysAlaAsnThrValAlaAspAsnSer	13372
3469 ACCGGCTTCCCGCGCGCGCGCGCGCGCG	35000
1338	ValAlaAsnAsnSerLysSerGluProLysSerArgLysG..ArgArg	13533
3501	GGATTGGCCGCAACCGCAG.....	3519
1353	gSerLieserGlnProGlnGluThrSerAlaGluGluThrThrAlaAs	13707
3520CCCAACCGCGA	3531
1370	eThrAspGluThrThrThrLeuAlaAspAsnSerLysArgSerLysProAsn	1386
3532	CCCCAACCGCAG.....	3543
1387	ArgArgSerArgArgSerValAlaArgSerGluProThrValThrAsnGlySe	1403
3544CGGACCTGATCGCCGTTATGCGCAATA	3571
1403	AspArgSerThrThrValAlaLeuAlaArgAspLeuThrSerThrAsnThrAsnA	1420
3572	GGCGTTTGAGTGAATTTTCCGCC.....ACGTCAAACGC	3606
1420	ValAlaLieserAspAlaMetAlaLysAlaGlnPheValAlaLeuAsnVal	1456
3607	GTTTTCGCCGTACAGGACGAATTTGGACGCGGTG...TTTGCCGAAGACGC	3653
1437	GlyLysAlaValSerGlnHisLieserGlnLeuGluMetAsnAsnGluG	1453
3654	CCCCAAGCGCGTTTGGACAAGGGGCGATCGGGACACCAACAGTACCGTT	3703
1453	gGlnTyrAsnValThrPvalSerAsnThrSerMetAsnLysnTyrSerS	1470
3704	CGCAAGATTTCCGGCCTACCGCCCAACAAACCGACCTGCGCCAAATCGGT	3753
1470	eSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly	1486
3754	ATGCAGAAAACCTGGGACGGCGG...CGCGCGGCATCCCGTTTTCGCA	3800
1487	ThrAspGlnThrThrLieserAsnAsnValGlnLeuGlyGlyValPheThrTyr	1503
3801	CAACCGCGACCGGAAACACCTTGAGACGACGGCATCGCAACTCGGACACGGC	3850
1503	ValAlaArgAsnSerSerAlaAsnPheAspRysAlaSerSerLysAsnThrLeuA	1520
3851	TTTCCCAACGGTCCGTTTTCGGGCAATPACGCGATCGGACGTTGACATC	3900
1520	LaGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisThrTyrLeu	1536
3901	GGCATAGCGCGGGCGCGGCTTGTAGTAAACGCGACGCTTTCGACAGCGAT	3950
1537	GlyThrAspLeuGlyTyrTrpLysPheGlnSerAsnLeuLysThrAsnHis	1553
3951	CACAGCGAAATACCGCGCGCGCTGCTGCATTAACGACGACAGAT	4000
1553	AsnAlaLysPheAlaArgHisThrAlaGlnPheThrLeuThrAlaGlyL	1570
4001	ACCGCGACGGTTTCGGCGGATTCGGCATCGCAACCGCACATCGGGCGAACG	4050
1570	ysaLarPheAsnLeuGlyAsnPheGlyThrThrProIleValGlyValArg	1586
4051	CGGTAATTTCGTCGCAAAAACGGGATTAACGATACGAAACGTAATATCGC	4100
1587	TyrSerTyrLeuSerSerAsnAlaAsnPheAlaLeuAlaLysAspArgLysLeu	1603
4101	CACCCCGGGGCTTGATTAACCGCTACCGCGCGGGCGATTTAAGCGACATGT	4150
1603	ysAlaAsnProIleSerValLysThrAlaPheAlaValAlaLysPheLeuT	1620

```

4151 ATTGATTCGAACCGCGGCACACATTCATCCAGCCCTTATTGAGCCTG 4200
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1620 yrlthryhlsleuglyglu...pheserValthrprolleuSerAla 1635
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
4201 TCCGAT...ACCGATGCCGCTTCGCGCAAGTCCGAAAGCGCGCATAC 4247
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1636 ArgTyrAspThrAsnGlnGlySerGlyLysile.....AsnValAsnG1 1650
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
4248 CGCGCATATGGCGGAGATTTCGGCAAAACCGGAGTGGGGAATGGGGCG 4297
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1650 nTyrAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyL 1667
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
4298 TAAAGCCGCAATCAAGGTTTCAAGCTGTCCCTCCAGCGTCCGCGCGC 4347
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1667 eulysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThr 1683
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
4348 AAGGGCGCGCAATTGGAAAGCGGACGACAGCGCGCGCATCAAAATTAGCCTA 4397
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1684 LysAlaLysGlnAlaGlyLysGlnLysThrAlaGluLeuLysLeuSerPh 1700
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
4398 C 4398
1700 e 1700
seq_name: /cgn2_6/prodata/1/laa/PCTUS_COMB.pep:PCT-US95-10661A-5

```

```

seq_documentation_block:
; Sequence 5, Application PC/TUS9510661A
; GENERAL INFORMATION:
; APPLICANT: Washington University, et al.
; TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10661A
; FILING DATE: 16-AUG-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: FP-59941/RFT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
PCT-US95-10661A-5

```

```

alignment_scores:
Quality: 1581.50      Length: 1867
Ratio: 1.677          Gaps: 48
Percent Similarity: 50.509      Percent Identity: 25.174

```

```

alignment_block:
us-09-303-518d-653 x PCT-US95-10661A-5
Align seg 1/1 to: PCT-US95-10661A-5 from: 1 to: 1702

```

```

64 CGCTTCTGCCCGCTTACTTATGACATATGCTGTGCTGGCATTCGTGCC 113
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
5 LysPheLysLeuAsnPhelAlaLeuThrValAlaLysLeuThrPr 21
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
114 CCAAGCGCGCGGCGGACACACTATTTCGGCATCCACTACCACTATCATC 163
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
21 oTyrThrGluAlaAlaLeuValAlaArgAspAspValAspIlePheA 38
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
164 GCGACTTGGCGGCAAAATTAAGCAAGTTGACGTGCGGCGCAAGATATT 213
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
38 rGAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
214 GAGCTTTAGCAACAAAAGGGAGTGTGGCGCAAAATCGACGAAAGC 263
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
55 GluValAlaArgAspLysAsnAsnHisSerLeuGlyAsnValLeuProAsnG1 71
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
264 C...CCGATGATGATTTTCTGTGATCG...CGTAACGGCGTGGCGG 307
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
308 CATTTGGCGGCGGATCAATATATTGTGAGCGTGACAT...AAGCGCGC 354
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
88 hrLeuLeuAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
355 TATACAAATGTTGATTTGGT.....GCGGAGCGAAGCA 389
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
390 TCCCGATCAGCACCGC.....TTTCTTACCAATG 421
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
121 nAspLysSerHisArgAspValSerSerGluGluAsnArgTyrPheSerV 138
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
422 TGAAGAATAATATTAATTAAGCAGGACTAACGC..... 456
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
138 alGluLysAsnGlnLysProThrLysLeuAsnGlyLysAlaValThrThr 154
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
457 .....CATCTTATGGCGGCGATTTATCATATCCGCGTTTGCA 494
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
155 GluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLeuAs 171
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
495 CAATTTGTCACGATGACAGACCGCTGTGAG...ATGACCACTTATATG 541
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
171 pLysPheValThrGluValAlaIleProIleGluAlaSerThrAlaSerSera 188
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
542 ATGGGTGGAATATACGCTGATTTAAATTAATACCGTATGCTGTGCAATC 591
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
188 spAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeu 204
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
592 GGAGCAGCGACACATATTGGCGGCTGTGATGAAGCAAA..... 630
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
205 GlySerGlyThrGlnPheIleTyrLysLysGlyAspAsnTyrSerLeuI1 221
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
631 .CCCAATTAACCGGCAAGATTCAATATCAT.....ATTGCAAGCG 667
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
221 eLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyAspA 238
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
668 CATATCTTGCTGCTGCTGCGTGGCAATATCCCTTGCAAAATGATGACGT 717
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
238 leryThrTyrGlyIleValGlyThrProTyrLysValAsnHisGluAsn 254
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
718 GGTGCGACAGTCACTTAGTAGCGCAAAATTAATTAACATAGC...CCATA 764
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
255 AsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspProLy 271
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
765 TGGTTTTTACCAACAGGA.....GGCATTTTGGGAGCA 799
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
271 sGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValIleuGlyAspS 288
||||| : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

800 GTGGCTACCAATGTTTATCTATGATGCCAAAGCAAAAGTGGTAAAT 849
      ||||| :|||:||||| :|||:||||| :|||:|||||
288 erglySerProleuPheValItyrAsnIySrglyLysTrpleuPhe 304
      ||| :||| :||| :||| :|||
850 AATGGGATTATGCAACAGCACCCCTATATAGAAAAGCAATGGCTT 899
      ||| :||| :||| :||| :|||
305 LeuGlySerTyraPheTrpAlaGlyTyraSnlYsLysSerTrpGlnG 321
      ||| :||| :||| :||| :|||
900 CCAGTACTTGTAAAGATTGGTCTATGATGAATCTTTGCTGGAGATA 949
      ||| :||| :||| :||| :|||
321 uTrpAsnIleTyrrLySproGluPheAlaLysThrValLeuAspLy 338
      ||| :||| :||| :||| :|||
950 CCCATTTCAGTATCTACGAACCATCATCAAAATGGAAATACTTTTAA 999
      ||| :||| :||| :||| :|||
338 hr..... 338
1000 GACATAATAATATGGCGCAGAAATAATGATCCAAACATAAATATTC 1049
      ||||| :|||:||||| :|||:||||| :|||:|||||
339 .....AlaGlySerLeuIleGlySerAsnThrGlnTyraS 350
1050 TCTACTATATGATTAAACACGACCGTT.....CAATTGT 1087
      :||| :||| :||| :||| :|||
350 nTrpAsnProThrGlyLysTrpSerValIleSerAsnGlySerGluSer 367
1088 TTAATGTTTCTTATCCGAGACAGCAAGAACCTGTTTATCATGCTGCA 1137
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
367 euAsnValAspLeuPheAspSerSerGlnAsp..... 377
1138 GGTGGGGTCACACAGTATTCAGCCACAGCTGATATATGAGAAATATTC 1187
      :||| :||| :||| :||| :|||
378 .....ThraSperLysLySAsnSnlHisGlySerValThr 390
1188 CTTTATGACAAAGAAAGTGAATTGATATCTACACGACATCATCACC 1237
      :||| :||| :||| :||| :|||
390 rLeu.....ArgGlySerGlyThrLeuThrLeuAsnAsnSnlLeuSpG 405
1238 AAGCGCGGGGGGTTTGTATTTTGAGGGTAAATTATTCAGGTC...TCCCT 1284
      ||||| :|||:||||| :|||:||||| :|||:|||||
405 InGlyAlaGlyGlyLeuPhePheGlnGlyAspTyrglyValLysGlyThr 421
1285 AAAACACACGAAACGTGGCAAGGGCGGGCTCATATCAGTATGGCAG 1334
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
422 SerAspSerThrThrTrpLysGlyAlaGlyValSerValAlaAspGly 438
1335 TACCGTACTTGGAAAGTAAACGGCGTGGCAACGACCGCTGGCCAAA 1384
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
438 sThrValThrTrpLysValHisAsnProLysSerAspArgLeuAlaLys 455
1385 TCGGCAAGGCACGCTGCTGGTTCAGCCAAAGGGGAAACCAAGGCTCG 1434
      ||||| :|||:||||| :|||:||||| :|||:|||||
455 leGlyLysGlyThrLeuIleValGlnGlyLysGlyLysAsnLysGlySer 471
1435 GTGCGCTGGCGGCGGTAAGTCATCTTGATCAGAGCGGACATCA 1484
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
472 LeuLysValGlyAspGlyThrValIleLeuLysGlnGlnAlaAspAla 488
1485 AGGCAAAAACCAAGCCCTTAGTGAATCGCTGTCAGCGGAGGAGGGA 1534
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
488 AsnLysValLysAlaPheSerGlnValIleValSerGlyArgSer 505
1535 CGGTGCACTGAATGCCGATTAATCAGTTCACCCGCAACAACCTATTC 1584
      ||||| :|||:||||| :|||:||||| :|||:|||||
505 hrValValLeuAsnAspAspLysGlnValAlaAspAsnSerIleTyrrPhe 521
1585 GGGTTCGGCGGAGAGCTTGGATTGACAGGGCATTCGTTTCGTTCCA 1634
      ||||| :|||:||||| :|||:||||| :|||:|||||
522 GlyPheTrpGlyGlyArgLeuAspAlaAsnGlyAsnLysLeuThrPheG 538
1635 CCGCATTCAAATATCCGATGAGGGGAGATGTCACACCAATCAATCAG 1684
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
538 uHisIleArgAsnIleAspAspLysAlaArgLeuValAsnHisAsnThrS 555

```

```

1685 ACAAGATCCACCCTTACCATTCAGCGCAATTAACATATTAATCT 1728
      :||| :|||:||||| :|||:||||| :|||:|||||
555 euLysThrSerThrValThrIleThrGlyLysLeuIleThrAspPro 571
1728 ..... 1728
572 AsnThrIleThrProTyraSnlLeaSpAlaProAspGluAspAsnPro 588
1728 ..... 1728
588 rAlaPheArgArgIleLysAspGlyGlyInLeuTyrrLeuAsnLeuGlu 605
1728 ..... 1728
605 snTyrrThrTyrrAlaLeuArgLysGlyAlaSerThrArgSerGluLeu 621
1729 .....ACAACCGGCAATAC..... 1743
622 ProLysAsnSerGlyLysSerAsnGluAsnTrpLeuTyrrMetGlyLys 638
1744 .....AACACTTGGATACCAAAAAG 1765
638 rSerAspAlaAlaLysArgAsnValMetAsnHisIleAsnAsnGluArg 655
1766 AATTCCTATCAACGGTGTGGCGAGAGAAGATGCAACCAACGAC 1815
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
655 euLysnGlyPheAsnGlyTyrrPheGlyGlyGlnGlyLys...LysAsnAsn 670
1816 GGGCGGCTCAATCTGAATTACCAACCGGAGAGGAGTGGCATTTACT 1865
      ||||| :|||:||||| :|||:||||| :|||:|||||
671 GlyAsnLeuAsnValThrPheLysGlyLysSerGlnGlnAsnArgPhe 687
1866 GCTTCCGGCGGACAAATTTAAAGCGCAATATCACGCAACAAACGCA 1915
      ||||| :|||:||||| :|||:||||| :|||:|||||
687 uLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuLysValGlyLysGly 704
1916 AACTGTTTTCAGCGGCGAGACCGACACGCGCTACATCATATTAGGA 1965
      ||||| :|||:||||| :|||:||||| :|||:|||||
704 hrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIleAlaGly 720
1966 AGCGGGGTCAAAAATGGAAGT.....ATCCACAGAGAAATCGT 2009
      :||| :||| :||| :||| :|||
721 IleSerSerThrLysLysAspGlnHisPheAlaGluAsnAsnGluVal 737
2010 GTGGGCAACAGATTGATGACCGCAGCATTTAAACGCAAACTTCATA 2059
      ||||| :|||:||||| :|||:||||| :|||:|||||
737 lValGluAspAspTrpIleAsnArgAsnPheLysAlaThrAsnIleAsn 754
2060 TTCAGGCGGCAACGCGGTGTTTCC...CGCAATGTGGCCAAAGTGGA 2106
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
754 alThrAsnAsnAlaThrLeuTyrrSerGlyArgAsnValAlaAsnIleThr 770
2107 GCGGATTTGCAATTAAAGCATCAGCCCAAGAGTTTTCGCTGGCACC 2156
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
771 SerAsnIleThrAlaSerAspAsnAlaLysValHisIleGlyTyrr.... 785
2157 GCATCAAAACCCACACATCTGTACACGTTGCGACTGGAGGGTCTGACA 2206
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
786 lLysAlaGlyAspThrValCysValAlaArgSerAspTyrrThrGlyTyrrAl 802
2207 GTTGTACGGAATAAACCATTTACCGAGATAAAGTATGCTCATTTAGGC 2256
      :|||:||||| :|||:||||| :|||:||||| :|||:|||||
802 hrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSerPheSnl 817
2257 AAGCCGACATCAGAGCAATGTACGCTTGCCGATCAGCTCATTTTAA 2306
      ||||| :|||:||||| :|||:||||| :|||:|||||
818 AlaThrAsnValSerGlyAsnValAsnLeuSer..... 828
2307 TCTCACAGACTTGGCACATCAACGGCAATCTTAGTCAGCGGAGACA 2356
828 ..... 828
2357 GCGACTATACGGTTACGGCAACGCCACCAAAAGCGCAACCTCAGCCTC 2406

```

828 828
2407 GTGGGCAATGCCCAAGCAATTTAATCAAGCCATTTAAACGGCAAC 2456
|||||.....:|||||
829 ...GlysnAlasnPhenValLeuGlyLysAlaAsnLeuPhenGlyThrI 844
2457 ATGGCTTCGGACATGCTTCATTTAATCTAAGCAACAGCCGTACAAA 2506
|||||.....:|||||
844 eSerGlyThrGlyAsnSerGlnValArgLeuThrGluAsn..... 857
2507 ACGGCACTGACGCTTCGACAGCAAGCTAAGCAACGTAAGCATTCC 2556
857 857
2557 GCACTCAAGCGCAATGTCTCCCTAGCCGATAGGCACTATTCATTGGA 2606
857 857
2607 AAACAGCCGCTTACCGGAAAAATCAGCGCGGCAAGATACGGCATTAC 2656
857 857
2657 ACTTAAAGACAGCAATGACGCTGCCGTGGGACAGCAATTAGCAAT 2706
|||||.....:|||||
858SerHisThrHisLeuThrGlyAspSerAsnValAsnGln 870
2707 TTAACCTTGACAACGCCACCATTACACTCAATTCGCGCTATGCAcCGA 2756
|||||.....:|||||
871 LeuAsnLeuAspLysGlyHisIleHisLeuAsnAla.....GlnsnAs 885
2757 TGGCGCAGCGCGCAACCGGCAAGTGGGCAAGATGCCCGCGCGCGCTT 2806
|||||.....:|||||
885 pAlaAsnLysValThrThr..... 891
2807 CGCGCGCTTCCTATTTATCCGTTACGCCGCCAATTGCGCAGATCCCGT 2856
891 891
2857 TTCACACGCTGACGCTAAACGGCAATTTGACGCTCAGGACATTCCG 2906
|||||.....:|||||
892 TyrAsnThrLeuThrValAsn...SerLeuSerGlyAsnGlySerPheTy 907
2907 CTTATGTCGGAACCTTCGCTACCGCCAGCGGCAATTTGAAGTCGCGG 2956
|||||.....:|||||
907 rTyLeuThrAspLeuSerAsnLysGlnGlyAspLysValValThrI 924
2957 AAAGTCCGAAGGCATTACACTTGGCTGTCAACAATACCGGCAACGAA 3006
|||||.....:|||||
924 ySerAlaThrGlyAsnPhenThrLeuGlnValAlaAspLysThrGlyGlu 940
3007 CCGGTAAAGTCCGAGCAATTTGAGGTAGTGAGAGAAAGCAACACACACC 3056
|||||.....:|||||
941 ProThrLys...AsnGluLeuThrLeuPhenAspLaserAsnAlaThr.. 955
3057 GCTGTCCGAAATCTTAATTTCAACCTGCACAAAGCAACAGCTGCATCCG 3106
|||||.....:|||||
956 ...ArgAsnAsnLeuAsnValSerLeuValGlyAsnThrValAsnLeu 971
3107 GCGCATGGCGTTATCAGCTTATCCGCAAGACGGCGAGTTCGCGCTCAT 3156
|||||.....:|||||
971 LysAlaTrpLysTyLysLeuArgAsnValAsnGlyArgTyAspLeuTy 987
3157 AATCCG..... 3162
888 AsnProGluValGluLysArgAsnGlnThrValAspThrThrAsnIleTh 1004
3162 3162
1004 rThrProAsnAsnIleGlnAlaAspValProSerValProSerAsnAsn 1021
3162 3162

1021 LuGluIleAlaArgValGluThrProValProProAlaProAlaThr 1037
3162 3162
1038 ProSerGluThrThrGluThrValAlaGluAsnSerLysGlnJusery 1054
3163GTCAAAGAACAGACCTTCCGACAAACTCGCAAGCGG 3202
|||||.....:|||||
1054 sThrValGluLysAsnGluGlnAspAlaThrGluThrThrAlaGlnAsn 1071
3203 GAGA..... 3207
1071 LysGluValAlaGluGluAlaLysProSerValLysAlaAsnThrGlnThr 1087
3208ACGA 3212
1088 AsnGluValAlaGlnSerGlySerGluThrGluGluThrGlnThrGln 1104
3213 GCGCGCTTGACGCA..... 3228
1104 uLleLysGluThrAlaLysValGluLysGluGluLysAlaLysValGlu 1121
3228 3228
1121 ySerGluLysAlaLysValGluLysAspGluIleGlnAlaProGln 1137
3229AACAGGCACACTTGGCGCCAA..... 3252
1138 MetAlaSerGluThrSerProLysGlnAlaLysProAlaProLysGluVal 1154
3252 3252
1154 LserThrAspThrLysValGluGluThrGlnValGlnAlaGlnProGln 1171
3252 3252
1171 hrGlnSerThrThrValAlaAlaGluAlaThrSerProAsnSerLys 1187
3253CAACAGCGCGGAAAAAGACAGCGCAAGCGCTTGA 3287
1188 ProAlaGluGluThrGlnProSerGluLysThrAsnAlaGluProValTh 1204
3288 CGCGCTGATTCGCGCGCGCGCAAT..... 3312
1204 rProValAlaSerLysAsnGlnThrGluAsnThrThrAspGlnProThr 1221
3313GCCACGAAAGCAAGGATGTTGCC 3339
1221 LuArgGluLysThrAlaLysValGluThrGluLysThrGlnProPro 1237
3340 GAACGCGCGCGCGAGGACGCG..... 3360
1238 GlnValAlaSerGlnAlaSerProLysGlnGluGlnSerGluThrValGln 1254
3361GGGAAAAATGCC...GGCATTTATGACGCGG 3388
1254 nProGlnAlaValLeuGluSerGluAsnValProThrValAsnAlaG 1271
3389 ACGAAGAGAAAAACGGGTGACGCGGATPAAGAC..... 3423
1271 LuGluValGlnAlaGlnLeuGlnThrGlnThrSerAlaThrValSerThr 1287
3424ACGCGCTT 3431
1288 LysGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrAlaI 1304
3432 GCGGAACAGCGGAGCGGAAACCGCGCGCTAC..... 3468
1304 eThrGluThrAlaGluLysSerAspLysProGlnThrGluThrAlaAla 1321
3468 3468
1321 eThrGluAspAlaSerGlnHisLysAlaAsnThrValAlaAspAsnSer 1337

369 TTTTGTCGCGAGGAACCATCCGATCCAGCACCGGCTTTCTTACCAA 418
118 naspyalalaspysgluasn.....glutylrargv 129
419 TTGTGAAAGAAATTAATTAAGCAGGACTAAAGCCATCCTTATGGC 468
129 alvalgluglnasnsntylglupro.....Hislyalatr 141
469 GGC.....GATTATCATATGCCGCTTGCA 494
142 G1ylalaseranleuglyargleuglasptryasnmetalaargpneas 158
495 CAAATTTGTCAGATGAGCAGACCTGTGAGATGACCATTAATGATG 544
158 nlyserhealthrgluvalalaproillealaprothralspalaiglyg 175
545 GGTGG...AAATACGTGATTTAAATAAATACCTGATCGTTGCAATC 591
175 lyleuasprthrtyltylaspysasnargpserphevalargile 191
592 CGAGAGGCAGACAA.....TATTTGGGCTGTGATGACGACGACCAA 635
192 G1yalaglyarggluleuvaltyrglulysgllyaltryhnglul 208
636 TAACCGCAAGATTCATATCATATTTGCAAGCGCATATTTGGCTGTCG 685
208 yasn...glulysgllytryaspneu..... 215
686 GTGCAATACCTTTGACCAAAATGATCAGGTGGCAGACGATCACTA 735
216ArgthrglulgluleuileglyrhegllyasnHisnlysgllytr 230
736 GSTACGAAATAATTAACATAGC...CCATATGCTTTTAAACCAACGG 782
231 Serialaglugluleuylsnglnalaleuuserglinspalaletyr 247
783 AGGCTCATTTGCGCAGCAGTGGCTGCCAATGTTATCTATGATGCA 832
247 rgllyallegllyaspserylserproleuphealpheasprysgl 264
833 AGCAAAATGTTATTAATGAGGATTCGAAACAGCAACCCCTATATA 882
264 ysasnlglttrpvalrphelengllythrtyrlytrpalaiglytr 280
883 GGAATAAGCAATGCTCCAGCTAGTTCGTAAGATGTTCTTATGATGA 932
281 lylslysserttprglnglutrpsanlletryllyslglnphealasp 297
933 AATCTTTGCTGGAGATACCATTCAGTATTCAGAACCCATCAAAATG 982
297 sile.....Lysglinhisaspasn. 303
983 GGAATACTTTTACGACATATATATGCGCAGCAAAATGATGCC 1032
304Alaiglythrvallysgly 309
1033 AAACATAACACTATCTCTACTTATGATTAATAACAGCAACCGTCA 1082
310 asnglygluhishis.....Tryltystrthr 318
1083 ATTGTTAATGTTCTTTATCCGACAGACAGAACCTGTTATCATG 1132
318 318
1133 CTGCGAGTGGGTCAACAGTTAT.....CGACCCAGACTG... 1167
319Glythrassnerhisileglyserthrhalavalargleuala 332
1168AATATGAGAAATATTTCTTATGACAA 1199
333 Asnasnglulgllyaspalalasnnglyglinsvalthrphenglualsp. 348

1200 AGCAAAAGGTGAATGTACTTACGACAAATCAACCAAGCGCGGCG 1249
349Asnglythrleuvalleuasnlnlsnnglinsgllyalaglyg 364
1250 GTTTGATTTTGAAGGTAATTTAGGCTCCGCTAAAAACAACAA... 1296
364 lyeupheglulysgllyasptrythrvallysglyalalasnasp 380
1297 AGGTGGCAAGCGCGGCGCTTCATATCAGTATGAGGAGATCCGTTACTG 1346
381 Thrtrpleuglyalaclytleasprvalalaspelylslyvalaitr 397
1347 GAAAGTAAACGGCGGTGCGCAACGCGCTGTCCAAATGCGCAAGCA 1396
397 rglvallylaspnproasnlylaspargleualaylsileglyl 414
1397 CGTGTGTTCAAGCCAAAGGGAACCAAGGCTCGGTGAGCGTGGC 1446
414 hrleuengluleasnnglythrlyvalalasnnglyglleuylsval 430
1447 GACGTAAGTCACTTATGATGACGAGCGGACGATCAAGCAAAACA 1496
431 Aspelythrvalilleuasnnglnlysalaspalaaspelylsval 447
1497 AGCCTTATGTAATCGGCTGGTGCAGCGGAGGAGCGATGCAATGA 1546
447 nalapheserglnvalglylvalserglylthrleuvalleu 464
1547 ATGCCGATATCACTTCAACCGCGCAACCTGATTTGGCTGGCGG 1596
464 snserferasnnglnlleasnproaspasnleuylrhegllyrheargly 480
1597 GGAAGTTGGATTTGAACGGGCAATTCGTTGCTCCAGCGATCAAA 1646
481 G1ylargleuaspalaasnnglylaspaleuylthrphenglinsile 497
1647 TACCGATGAAGGGGATGATGTTGTCACCAATCAAGCAAAAGATCA 1696
497 nvalasprgluglyalarglilevalasnHisntrpaspnlsalaser 514
1697 CGTTACCATTTACGCGCAATTAAGATTAATGATCAACCGGCAAT 1740
514 hrlethrleuthrnglyllylserleuilethrassnproasnserleu 530
1741AACAACACTTGATACAA..... 1761
531 ValHisserlelelinsnsptrylaspgluasprtylserlytyr 547
1762AAGAAATTCGCTAC..... 1776
547 rargproargargproleproglngllylspaleuyltyrlyl 564
1776 1776
564 ylrargtyrtyrallaleuylseryglyllyalgluasnalaprometro 580
1776 1776
581 Gluasnglyvalalagluasnaspnasptryllepheleglytyrthr 597
1776 1776
597 ngluglualaarglylaspasnalametlanshslsypasnaspargyleg 614
1777AAGGTTGTTTGGCGAAGAAAGATCAACCAAAAGCAAGCGG 1818
614 lypspgluglyllyrhepneaspelugluasnnglyllysgllyhnsngly 630
1819 CGGCTCAATGATTAATCAACCGGAGAGCGGATCGCATTTACTGCT 1868
631 AlaleuasnleuasnphasnnglyllysserAlaglaasnargpheleu 647
1869 TTCGGCGGCAACAATTTAAACGGCAATATCACGCAACCAACGGCAAC 1918

647	uThrgIyGIyAlaAsnLeuAsnGlyLysIleSerValThrGlnGlyAsnV	664
1919	TGTTTTCAGGGGCGAGACCGACACCGCATTCATCAATTCATTAGAACG	1968
664	aIleuLeuSerGlyArgProHisAlaArgAspPheValAsnLys	680
1969	GGGTGCTCAAAAATGGAAGT.....ATCCACAAAGAGAAATCGGTG	2012
681	SerSerIleAlaArgLysAspAlaHisPheSerLysAsnAsnLysValAlaP	697
2013	GGACACAGATTGGATGATCGCGACATTTAAAGCGGAAACTCCATTT	2061
697	eGlnAspAspTrpIleAsnArgTrpPheLysAlaAlaGluIleAlaValA	714
2062	..CAGGGCGGCAAGCGGTGGTTTCCCGCAATGTTCGCAAGTGGAAAGC	2109
714	sngInSerAlaSerPheSerSerGlyArgAsnValAspAspIleThrAla	730
2110	GATTGGCATTTTAAAGCAATCACGGCCACAGAGTTTTCGTGCGACCGCA	2159
731	AsnIleThrAlaThrAspAsnAlaLysValAsnLeuGlyTyr.....L	745
2160	TCAAGCCACACAATCTGTACAGCTTGGACGTGACGGGTGTGACAGATT	2209
745	sAsnGlyAspGluValCysValAlaArgSerAspTyrThrGlyTyrValTh	762
2210	GTACCGAAAAACCATTACCGGATTAAGTGAATGGTTCATTTGACCAAG	2259
762	YAsnThrArgLysLeuSerAspLysAlaLeuAsnSerPheAsp***Ala	778
2260	ACCGACATCAGAGGCAATGTACAGCTTGGCATCAGCTCATTTAAATCT	2309
779	ThrLysIleAsnGlyAsnValAsnLeuAsnGlnAsnAlaLeuValLe	795
2310	CACAGGACTTGCACACTCAAGCGCAATCTTACTGCAGGGGAGACACGC	2359
795	u...GlyLysAlaIleAlaLeuTrpGlyLysIleGlnGly.....	806
2360	ACTATACGGTTACGGCGCACGCCACCCAAAACGCAACTCAGCTCTGTG	2409
807Gln	807
2410	GGCAATGCCCAAGCAACATTATTCAAAGCCACATTTAAGGCGACACATC	2459
808	GlyAsnSerArgValSerLeuAsnGln.....	816
2460	GCGTTGCGACAATGCTTCAATTATCTAAGCAACAAGCCGTACAAACG	2509
816	816
2510	GCACTGTGACGCTTTCGACACACGCTAAGGCAAAAGCTAAGCCATTCCGA	2559
816	816
2560	CTCAAGCGCAATGTCTCCCTAGCCGATAAGGCGATTCATTTGAAA	2609
816	816
2610	CAGCCGCTTACCGGAAAAATCAGCGCGGCAAGGATACGGCATTCACCT	2659
816	816
2660	TAAAGACAGCAATGAGCGTCCGTGGCGACAGCAATTAGCAATTTA	2709
817HisSerLysTrpHisLeuThrGlyAspSerGlnValHisAsnLeu	831
2710	AACCTTGACAAACGCCACATTACACTCAATTCGCGCTATGACAGCATGC	2759
832	SerLeuAlaAspSerHisIleHisLeuAsnAla.....SerAspAl	846
2760	GGCAGGGCGCCCAACCGGCGACAGTGGCGCAATGCGCGCGCGCTTCCG	2809
	
846	agInSerAla.....	849
2810	GCGGTTCCTATTATCCGTTACCGCGCAACTTCGGCAAAATCCCGTTTC	2859
850AsnLysTyr	852
2860	AACACGCTGACGCTAAACGGCAATTTGAACGCTGACGGACATTCGCTT	2909
853	HisThrIleLysIleAsn...HisLeuSerGlyAsnGlnHisPheHisT	868
2910	TATGTCCGAACCTTTCGGCTACCGCGGCAAGCAATTTGAAGCTGGCGAA	2959
868	rLeuThrAspLeuAlaLysAsnLeuGlyAspLysValLeuValLysGluA	885
2960	GTTCCGAAGGCACTTACACCTTGGCTGTCCAAACATTCGCGACAGAAC	3009
885	LaSerAlaGlyHisTyrGlnLeuHisValGlnAsnLysThrGlyGluPro	901
3010	GTAAGTCTGACCAATTCGCGTACTGTGAGAGAAAGCAACACACCGCT	3059
902	AsnGln.....GluGlyLeuAspLeuPheAspAl	911
3060	GTCC.....GAAATCTTAATTTACCCCTCGCAAAACGAC	3094
911	aSerSerValGlnAspArgSerArgLeuPheValSerLeuAlaHisnHis	928
3095	ACGTGATCCGCGCCGCGCATTCGCTTATCGACTTATCCGAAGACGGCGAC	3144
928	YValAspLeuGlyAlaLeuArgTyrThrIleLysTrpHisAsnGlyLe	944
3145	TTCCGCTGCAATATCCGCTCAAGAAACAAAGAGCTTTCGCAAACTCGG	3194
945	ThrArgLeuTyrAsnProTyrAlaGly.....AsnGlyArgProVa	958
3195	CAGCGCGGAGAAACAGAGCCCGCTTACGCGCAAAACAGGACACACTTG	3244
958	IlyProAlaProSerProAlaAlaAsnThrAlaSerGlnAlaGlnLysA	975
3245	CCGCC.....	3249
975	LaThrGlnThrAspGlyAlaGlnIleAlaLysProGlnAsnIleValAl	991
3250AAACACAGCG	3260
992	AlaProProSerProGlnAlaAsnGlnAlaGluGluLeuLeuArgGlnGlnAl	1008
3261	GGAAAAAGACAACGGC.....CAAGCGCTGACGCGCTGATTTCGG	3301
1008	aLysAlaGluGlnValLysArgGlnGlnAlaAlaGluAlaLysValA	1025
1025	LaArgGlnLysAspGluGlnLysArgLysAlaAlaGluIleAlaArg	1041
3352	CAG.....GCAGCGGGGGAATAATCCGCGCATTTAGCAGCG	3386
1042	GlnGlnGlnGluAlaArgLysAlaAlaGluLeuAlaAlaLysGlnLysAl	1058
3387	GGAGAGACAGAAAAACGGGTGCAGAGGGGATAAAGACACCGCCTTGGCGA	3436
1058	agLysAlaGluLysArgLysAlaArgLys.....LeuAla	1069
3437	AACAGCGGGAACGGGAACCGCGCGGCTACACCGCTTCCCGCGCG	3486
1069	rgGlnLysAlaGluGlnLysAlaSerHisGlnAlaAsnAlaLysProLysArg	1085
3487	CGCGCGCGCGCGGATTTCCCGCAACG.....	3516
1086	ArgArgArgArgAlaIleLeuProArgProProAlaProValPheSerLe	1102
3516	3516
1102	uAspAspTyrAspAlaLysAspAsnSerGlySerIleGlyAsnLeuA	1119

```

3516 ..... 3516
1119 laargvalprolleargmetglarggluleulleasnasyrglglu 1135
3516 ..... 3516
1136 lleproleuglgluleuglaspdluagluglgluararglinal 1152
3516 ..... 3516
1152 athrghlnpheniserrysserargasnargala1leserterglup 1169
3516 ..... 3516
1169 roseraseraspdluasplaserterglu1serthrserysphyhis 1185
3516 ..... 3516
1186 proglinsprasnthrghluleuHlglulysvalgluthralagluleugl 1202
3517 ..... CAGCCCAACCCGCAACCCGACGCGACCTGA 3553
1202 nproarglalaaglproargthrghlnalaalaglinalasp1av 1219
3554 TCAGCCGTTATGCAATGAGCGTTGAGTGAATTTCCGCGACGCTGAC 3603
1219 alserthrsnthraserlaserlaseraspalamelalaserthrgln 1235
3604 AGCGTT.....TTCGCGCTACAGAGCAATTTGACCG 3635
1236 SerlleuleuaspthrghlYalatyLeuthrarghislleaglinal 1252
3636 CGTGTTCGCCAAGACCCGCCAGCGCTTGACCAAGCGGATCCGGG 3685
1252 sserarglaspalaglulysasnserValtrpmetelulasnthrghlY 1269
3686 ACACCAACATACCGCTTCCGCAATTTCCGCGCTACCGCAACAAAC 3735
1269 yrglYargaspYrAlaserlaglIntrYarglpheserterlyarg 1285
3736 GACCTGCGCAATCGGTATGCAAGAAAACCTGCGGACGGG...CGCGT 3782
1286 ThrghlnthrghlnleghlYleaspargserleuSerghluasmetghln 1302-
3783 CGGATCTCTGTTTCGCAACCGGACCGGAAACACCTTCAGCAGCGCA 3832
1302 eglYglYalValleuthrYraseraspserghlnHlstrpheasplnala 1318
3833 TCAGCACTCGGACGCTTCGCCACGCGCTTTCGCGCAATTCAGCGC 3882
1319 ..glYglYlYasnthrphelValghlnalaasnleuthrYglYlYtYr 1334
3883 ATCGGC...AGGTTGCACATCGGATCGGACGGGCGGCGGTTTAAAG 3929
1335 leuasnaspalatrprYValaglYasplleghlYasplleghlYserleuAr 1351
3930 CGGCGCTTTCAGACGCGCATCAGAGCAAAATCCGCGCGCTGCTGC 3979
1351 gserargleuInthrghlnlYsalaaSnphasnargthrsertleeg 1368
3980 ATTACGCGATTCAGGACAGATACGCGCAGGTTTCGCGGCAATTCGCGC 4029
1368 InthrghlYleuthrleuglYasnthrleuYslleasnghlnpheglulle 1384
4030 GAACCGCATCGCGCAACCGGCTATTTCTCCAAAAAGGATTCACG 4079
1385 ValProseralaglYlleargYraserargleuSerterlaasprYrly 1401
4080 ATAGCAAAAGCTCAATATCGCACCGCGCTTCGATTCACCGCTAAC 4129
1401 sleuglYaspsaserValYlYsValserSerMetlavalYlYstrhleut 1418

```

```

4130 GCGCGGCAATTAAGCAGATATTCATTCAAACCGCGCAACACATTTCC 4179
1418 hrleaglYleuasnphelatrYargphelys...ValglYasnleuthr 1433
4180 ATACGCGCTTAATTCGACCTGTCTATACCATGCGCTTCGCGCAAGT 4229
1434 ValYlYsProleuSerlaserlala1atYrpe...AlaasnYrglYlYsGl 1449
4230 CGGACGCGCGCAATACCGCGCTATTCGCGCAGATTTTCGCAAAACC 4279
1449 YglYalasnValglYlYsSerphelatrYrlysalaspsanglng 1466
4280 GCAGTCGGAATGGGCGTAAACGCCGAATCAAGTTTCACGCTGTC 4329
1466 lnglYrseralaglYalalaleuLeuthrYrargasnValthrleuasn 1482
4330 CTCACGCTGCGCGCGCAAGGCGCGCAATTCGAGCGCACACAGCGC 4379
1483 ValasnYlYserlYlethrYsglYlYsghlnleuglYlYsYsergl 1499
4380 GCGCATCAATTAAGCGTACCGCTGG 4404
1499 YglInlYlYslleghlnlleargPhe 1507

```

seq_name: /cgn2_6/plodata/1/1aa/6B_COMP.pep:US-08-296-791-6

seq_documentation_block:

```

; Sequence 6, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Gene III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1848 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; US-08-296-791-6

```

```

alignment_scores:
  Quality: 1513.00      Length: 1916
  Ratio: 1.519          Gaps: 60
  Percent Similarity: 51.983  Percent Identity: 25.418
alignment_block:

```

US-09-303-518D-653 x US-08-296-791-6 ..

Align seg 1/1 to: US-08-296-791-6 from: 1 to: 1848

```
64 CGCTTCGCGCCGCTACTTACCATATGCTGCTGCTGCGCATTCGCC 113
   :::::::::::::::::::::
5  LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21
   :::::::::::::::::::::
114 CCAAGCCGCGGCGGACACACTTATTTGCGCATCACTACCAATACATTC 163
   | :::::::::::::::::::::
21 oTyrThrGluAlaAlaLeuValAlaArgAspAspValAspTyrGlnIlePheA 38
   :::::::::::::::::::::
164 GCGACTTTCGCGAATAAAGCAAGTTTGCGAGCGGCGGCAAGATATT 213
   | :::::::::::::::::::::
38 rGAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54
   :::::::::::::::::::::
214 GAGGTTTACACAAAAAGGAGTGTGTGCGCAATCGATGACGAAGC 263
   | :::::::::::::::::::::
55 GluValArgAspLysLysAsnGlnSerLeuGlySerAlaLeuProAsnGly 71
   :::::::::::::::::::::
264 C...CCGATGATGATTTTCTGTGATGCG...CGTACGCGCGGCGG 307
   | :::::::::::::::::::::
71 yIlePheMetIleAspPheSerValAlaAspValAspLysArgIleAla 88
   :::::::::::::::::::::
308 CATTGCGGCGGCGATCAATATATGTGTAGCGTGCGCAT...AACGCGCG 354
   :::::::::::::::::::::
88 hrLeuValAsnProGlnTyrValAlaGlyValLysHisValSerAsnGly 104
   :::::::::::::::::::::
355 TATAACAATGTGATTTTGGT.....GCGGAGGGAAGCAA 389
   :::::::::::::::::::::
105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121
   :::::::::::::::::::::
390 TCCCGATGACGACCGC.....TTTCTTACCAAAATTG 421
   | :::::::::::::::::::::
121 nAlaLysSerHisArgAspValSerSerGluAsnArgTyrTyrThrV 138
   :::::::::::::::::::::
422 TGAAGAAATAATTAT.....AAGCA 444
   | :::::::::::::::::::::
138 aLgIuLysAsnAsnPheProThrGluAsnValThrSerPheThrLysGlu 154
   :::::::::::::::::::::
445 GGCACATACGCGCATCCTTATGCGCGATATATATGCGCGGTTGCA 494
   :::::::::::::::::::::
155 GluGlnAspAlaGlnLysArgArgGluAspTyrTyrMetProAlaGlyLeu 171
   :::::::::::::::::::::
495 CAAATTTGTCACAGATGACAGACCTGTTGAG...ATGACAGATTATATG 541
   | :::::::::::::::::::::
171 pLysPheValThrGluValAlaProIleGluAlaSerThrAlaAsnAsnA 188
   :::::::::::::::::::::
542 ATGGGTGGAATACGCTGATTAAATAAATACCTGATCGTGTTCGATC 591
   :::::::::::::::::::::
188 snLysGlyGluTyrAsnAsnSerAspLysTyrProAlaPheValArgLeu 204
   :::::::::::::::::::::
592 GGAGCAGCAGACAAATAT.....TG 611
   | :::::::::::::::::::::
205 GlySerGlyThrGlnPheIleTyrLysGlySerArgTyrGlnLeuI 221
   :::::::::::::::::::::
612 GCGGTGTGTAAGACGACCCATTAACCGGAAAGTTCATATCAT... 657
   :::::::::::::::::::::
221 eLeuThrGluLysAspLysGlnGlyAsnLeuLeuArgAsnTrpAspValG 238
   :::::::::::::::::::::
658 .....ATGGCAAGCGCATATCTTGGCTGCTGCTGCTGCTGCTGCT 687
   :::::::::::::::::::::
238 LysGluAspAsnLeuGluLeuValGlyAsnAlaTyrThrTyrGlyIleAla 254
   :::::::::::::::::::::
688 GGCATATACCTTTGGACAAATGATGATGATGATGATGATGATGATGATG 737
   | :::::::::::::::::::::
255 GLyThrProTyrLysValAsnHisGluAsnAsnGlyLeuIleGlyPheGly 271
   :::::::::::::::::::::
738 TAGGCAAAAAATTAAACATAGC...CCATATGGTTTATTAACAACAGGA 793
   | :::::::::::::::::::::
271 yAsnSerLysGluGlnLysSerAspProLysGlyIleLeuSerGlnAspR 288
   :::::::::::::::::::::
784 .....GGCTCATTTGGCGACAGTGCGTACCAAAATGTTTATC 819
   :::::::::::::::::::::
288 rOleuThrAsnTyrAlaValLeuGlyAspSerGlySerProLeuPheVal 304
   :::::::::::::::::::::
820 TATGATGCCCAAAAAAGCAAGTGTAAATTAATGGGATTTGCAACAGG 869
   | :::::::::::::::::::::
305 TyrAspArgGluLysGlyLysTrpLeuPheLeuGlySerTyrAspPheTr 321
   :::::::::::::::::::::
870 CAACCCCTATATGCAAAAAAGCAATGCTTCCAGCTAGTTCGTAAGATT 919
   | :::::::::::::::::::::
321 pAlaGlyTyr.....AsnLysLysSerT 329
   :::::::::::::::::::::
920 GGTTCATGATGAATCTTTGCGAGATACCATTCAGTATTCATGCAAA 969
   | :::::::::::::::::::::
329 rPcGlnLysTrpAsnIleTyrLysHisGluPheAlaGlyLysIleTyrGln 345
   :::::::::::::::::::::
970 CCAATCAAAAAATGGCAATATCTTTTAAACACAAATTAATATGCGGACG 1019
   | :::::::::::::::::::::
346 GlnTyr.....SerAlaGly 350
   :::::::::::::::::::::
1020 AAAAATCGATGCCAAATAAACATATCTCTACCTTATGATTAATAAA 1069
   | :::::::::::::::::::::
350 ySerLeuIleGlySerAsnThrGlnTyrThrPcGlnAlaThrGlySerT 367
   :::::::::::::::::::::
1070 CACGAACCGTT.....CAATTGTTTAATGTTTCTTATACGAG 1107
   | :::::::::::::::::::::
367 hrSerThrIleThrGlyGlyGlyGluProLeuSerValAspLeuThrAsp 383
   :::::::::::::::::::::
1108 ACAGCAGAGAACCTGTTATATCATGCTGCAGGTGGGGGTACACATTAACG 1157
   :::::::::::::::::::::
384 GlyLysAspLysPro..... 388
   :::::::::::::::::::::
1158 ACCCAGACTGAATTAATGAGAAAAATTTCTTTATGCAAAAGGAAGA 1207
   | :::::::::::::::::::::
389 .....AsnHisGlyLysSerIleThrLeu.....LysGlySerG 400
   :::::::::::::::::::::
1208 GTCAATTGATACTTACACAGCAATCAACCAAGGCGGCGGCTTGTGAT 1257
   | :::::::::::::::::::::
400 lYThrLeuThrLeuAsnAsnHisIleAspGlnGlyAlaGlyLysLeuPhe 416
   :::::::::::::::::::::
1258 TTTGAGCGTAATTTTACGGTC...TGCCTAAACAAACAAACGCTGCTG 1304
   | :::::::::::::::::::::
417 PheGluGlyAspTyrGluValLysGlyThrSerAspSerThrThrTyrLys 433
   :::::::::::::::::::::
1305 AGCGCGGCGGCTTCATATCATGATGATGATGATGATGATGATGATGATG 1354
   | :::::::::::::::::::::
433 sGlyAlaGlyValSerValAlaAspGlyLysThrValThrTyrLysValH 450
   :::::::::::::::::::::
1355 ACGGCGTGGCAACGACGCGCTGTCAAATGCGAAAGCAGCGTGTG 1404
   :::::::::::::::::::::
450 lAsnProLysTyrAspArgLeuAlaLysIleGlyLysGlyThrLeuVal 466
   :::::::::::::::::::::
1405 GTTCAAGCCAAAGGGAACCAAGGCTGCGTACGCGGCGGAGCGGTAA 1454
   | :::::::::::::::::::::
467 ValGluGlyLysGlyLysAsnGlnGlyLeuLeuLysValGlyAspGlyTh 483
   :::::::::::::::::::::
1455 AGTCATCTTAGATCAGCAGCGGAGATCAAGCAAAAAACAAGCCTTTA 1504
   | :::::::::::::::::::::
483 rValIleLeuLysGlnLysAlaAspAlaAsnAsnLysValGlnAlaPheS 500
   :::::::::::::::::::::
1505 GTGAAATCGCGTTGTCAGCGGCGGAGGCGGCTGCAACTGATCCGAT 1554
   | :::::::::::::::::::::
500 eGcIuValGlyIleValSerGlyArgSerThrLeuValLeuAsnAspAsp 516
   :::::::::::::::::::::
1555 AATCAGTTTCAACCCGACCAACTATATTTGCGCTTCCGCGGCGAGCTTT 1604
   | :::::::::::::::::::::
517 LysGlnValAspProAsnSerIleTyrPheGlyPheAlaGlyGlyAlaPle 533
   :::::::::::::::::::::
1605 GGAATTTGAACGGGATTCGCTTTCGTCACCGCATTCAAATACAGATG 1654
   | :::::::::::::::::::::
533 uAspLeuAsnGlyAsnSerLeuThrPheAspHisIleArgAsnIleAspA 550
   :::::::::::::::::::::
1655 AAGGGCGGATGTGTCAACCAACATCAAGACAAAGAAATCCACGTTTAC 1704
   :::::::::::::::::::::
```

550 spgIyAlaArgValValAsnHisAsnMetThrAsnThrSerAsnIleThr 566
1705 ATPACAGCAATAAGATATTACTACA.....ACCGCAATAA 1742
|||||..... |||||
567 IleThrGlySerLeuIleThrAsnProAsnThrIleThrSerTyrAs 583
1743 C..... 1743
583 nIleGluAlaGlnAspAspAspHisProLeuArgIleArgSerIlePro 600
1743 1743
600 yTrArgIleuTyrPheAsnGlnAspAsnArgSerTyrTyrThrLeuLys 616
1743 1743
617 LysGlyAlaSerThrArgSerGluLeuProGlnAsnSerGlyGluSerAs 633
1743 1743
633 nGluAsnTripleuTyrMetGlyArgThrSerAspAlaAlaLysArgAsnV 650
1744AACACTTGATAGCAAAAAAGAAATGGCTTCAACGGTTGGTTT 1788
|||||..... |||||
650 alMetAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPhe 666
1789 GGGGAGAAAGATGCACCAAAACG...AACGGGGGCTCAATCGATTA 1835
|||||..... |||||
667 GlyGluGluGluThrLysAlaThrGlnAsnGlyLysLeuAsnValIleThr 683
1836 CCAACCGGAAGAGCGGATCGCACTTACTCTTCCGGGGGCAAAATTT 1885
|||||..... |||||
683 eaSnGlyLysSerAspGlnAsnArgPheLeuLeuThrGlyGlyThrAsnL 700
700 euAsnGlyAspLeuAsnValGlyLysGlyThrLeuPheLeuSerGlyArg 716
1886 TAAACGGCAATATCAGCGCAAAACGCAACCTTTTCAGCGGCGAGA 1935
|||||..... |||||
700 euAsnGlyAspLeuAsnValGlyLysGlyThrLeuPheLeuSerGlyArg 716
1936 CCGACACGCGACGCTCAATCATTTAGGAAGCGGGTGTCAAAATGGA 1985
|||||..... |||||
717 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrLysLysAs 733
1986 AGGT.....ATCCACAGAGAGAATCGTGGGACACAGATGTGATCG 2029
|||||..... |||||
733 pProHisPheThrGlnAsnAsnGluValValIleGluAspAspIleAla 750
2030 ACCGCACATTTAAAGCGGAAACTTCATATTCAGGGCGGCAACGGCTG 2079
|||||..... |||||
750 snArgAsnPheLysAlaThrThrMetAsnValIleThrGlyAsnAlaSerLeu 766
2080 GTTTC...CGCAATGTTGCCAAGTGAAGGCGATTTGCATTTAACCA 2126
|||||..... |||||
767 TyrSerGlyArgAsnValAlaAsnIleThrSerAsnIleThrAlaSerAs 783
2127 TCACGCCCAAGCAGTTTGGGTGTCGACCGCATCAAGCCACACAATCT 2176
|||||..... |||||
783 nasAlaGlnValHisIleGlyTyr...LysThrGlyAspThrValC 798
2177 GTACAGCTTGGAGCTGACGGCTGTACAGTGTACCGAAACCAAT 2226
|||||..... |||||
798 ysvAlaArgSerAspTyrThrGlyTyrValIleThrCysHisAsnSerAsnLeu 814
2227 ACCACAGATAAGTGAATGCTTCATTTAGCAAGACGACATCAGAGCAA 2276
|||||..... |||||
815 Ser...GluLysAlaLeuAsnSerPheAsnProThrAsnLeuArgGlyAs 830
2277 TGTGACCTTCCGATCAGCTCATTTAAATCTCAGAGACTTCCACAC 2336
|||||..... |||||
830 nValAsnLeuThrGluAsnAlaSerPheThrLeu...GlyLysAlaAsnL 846
2327 TCAACGGCAATCTTAGTCAGCGGA..... 2352
|||||..... |||||
846 eupHeGlyThrIleGlnSerIleGlyThrSerGlnValAsnLeuLysGlu 862

2353 GACACGCACTATACGTTACGGCAACGCCAC... 2385
|||||..... |||||
863 AsnSerHisThrPheIleuThrGlyAsnSerAsnValAsnGlnLeuAsnLe 879
2386 .CAAAACGGCAACCTCAGCTCTGTGGCAATGCCAA..... 2421
|||||..... |||||
879 uThrAsnGlyHisIleHisLeu...AsnAlaGlnAsnAspAlaAsnL 894
2422GCACATTTAATCAGCCACATTAAGCGCAACATCGGCTCG 2466
|||||..... |||||
894 ysvAlaThrThrTyrAsnThrIleuThrValAsn.....SerLeuSer 907
2467 GACAAATGCTTCATTT.....AATCTAAGCAA 2492
|||||..... |||||
908 GlyAsnGlySerPheTyrTyrThrValAspPheThrAsnAsnLysSerAs 924
2493 CAACGCCGTACAAACGCCAGT..... 2514
|||||..... |||||
924 nLysValValValAsnLysSerAlaThrGlyAsnPheThrLeuGlnValA 941
2515CTGACGCTTCCGACAA 2532
941 laAspLysThrGlyGluProAsnHisAsnGluLeuThrLeuPheAspAla 957
2533 GCTAAGCA.....AACGTAAGCATTCGCGCATCAACGGCAATGT 2573
|||||..... |||||
958 SerAsnAlaThrArgAsnAsnLeuGluValIleThrLeuAlaAsnGlySerVa 974
2574 CTCCCTAGCCGATTAAGCGAGTATTCATTTGAAACACGCCCTTTACCG 2623
|||||..... |||||
974 L.....AspArgGlyAlaIleThrLysTyrLysLeuArgAsnValAsnG 988
2624 GAAAAATCAGCGCGGCAAGGATACGCGCATTCACCTTAAACAGACGCAA 2673
|||||..... |||||
988 LArgTyrAsp.....LeuTyrAsnProGlu 996
2674 TGGACGCTCCGCTCGGCGACGGAATTTAGCAATTTAAACGTTGACACGC 2723
|||||..... |||||
997ValGluLysArgAsnGlnThrValAspThrTh 1007
2724 CACCATTAACATTCATTCGCGCTATCGACACAGATCGCGCGCGCAA 2773
|||||..... |||||
1007 rAsnIleThrThrProAsnAspIleGlnAlaAspAlaProSerAlaGln 1024
2774 CCGGCAGT..... 2781
1024 eTrAsnAsnGluGluIleAlaArgValGluThrProValProProAla 1040
2782GCGGCAGATGGCGCGCGCGCGCT... 2805
1041 ProAlaThrGluSerAlaIleAlaSerGluGlnProGluThrArgProAl 1057
2805 2805
1057 aGluThrAlaGlnProAlaMetCgluGluThrAsnThrAlaAsnSerThr 1074
2806TCGGGC 2811
1074 LuThrAlaProLysSerAspThrAlaThrGlnThrGluAsnProAsnSer 1090
2812 CGTTCCTATTAATCGGTACG..... 2832
1091 GluSerValProSerGluThrThrGluLysValAlaGluAsnProProGlu 1107
2832 2832
1107 nGluAsnGluThrValAlaLysAsnGluGluGluAlaThrGluProThrP 1124
2833CCGCCAATTCGGCGAGA 2850
1124 roGlnAsnGlyGluValAlaLysGluAspGlnProThrValGluLysAsn 1140

2851	TCGCCGTTTAAACACGCGTGAACGGTAAAC..GGCAATATGAACGGCGACGG	2897
1141	ThnGlnThrAsnGluAlaThrGlnSerGluGlyThrGlnGluThrGln	1157
2898	AAcATTCCGCTTATGATCGGAA.....	2919
1157	nThrAlaGluThrLysSerGluProThrGlnSerValThrValSerGluA	1174
2920CTGTCGGC	2928
1174	snGlnProGluLysThrValSerGlnSerThrGluAspLysValValVal	1190
2929	TACCGAGCGCGCAATGAAGCTG.....	2952
1191	GluLysGluGluLysAlaLysValGluThrGlnGluThrGlnLysAlaPr	1207
2952	2952
1207	oGlnValThrSerLysGluProProLysGlnAlaGluProAlaProGluG	1224
2953GCGGAAGTTCCGAAGCACT	2973
1224	luValProThrAspThrAsnAlaGluGluAlaGlnAlaLeuGlnGlnThr	1240
2974	TACACTTGGCGTC.....	2988
1241	GlnProThrThrValAlaAlaAlaGluThrThrSerProAsnSerLysPr	1257
2988	2988
1257	oAlaGluGluThrGlnGlnProSerGluLysThrAsnAlaGluProValT	1274
2989AACATACCGGCAACGAGCCCGTAAGTCTCGAGCA	3024
1274	hrProValValSerGluAsnThrAlaThrGlnProThrGluThrGlnGlu	1290
3025	TTGACGCGTGGAGAAAGAAACAACACACACCACCGCTGCCGAAATCTTAA	3074
1291	ThrAlaLysValGlu..LysGluLysThrGlnGluValaProGlnVal....	1305
3075	TTTCACCCCTGCAAAAGCAACAGCAGTGCAGTCCGCGCATGGGCTTATCGC	3124
1306AlaSerGlnGluSerPro..LysGlnGluGlnProAlaAl	1318
3125	TTATCCGCAAAAGACGCGAGTTCGCCCTGCATTAATCCGTCGCAAGACAA	3174
1318	AlysProGlnAlaGlnThrLysProGlnAla..GluProAlaArgLysAsn	1334
3175	GAGCTTCCGACAAA.....	3189
1335	ValLeuThrThrLysAsnValGlyLysGluProGlnProGlnAlaGlnProGln	1351
3190CTCGGCAAGCGGGGAGAAACAAGGCC..GCCT	3220
1351	nThrGlnSerThrAlaValaProThrThrGlyGluThrAlaAsnSerTr	1368
3221	TGACGCAAAACAGCAGCACACTTGCAGCAAAACAAGCGCGAA.....	3264
1368	ysrProAlaAlaLysProGlnAlaGlnAlaLysProGlnThrGluProAla	1384
3265AAAGCAACCGCGCAACCTTTGA	3287
1385	ArgGluAsnValSerThrValAsnThrLysGluLysGluProGlnSerGlnThrSe	1401
3288	CGGCGTGATGGGGCGGGCGCAATGCCAGCAAGCAAAAGGCNAAGTGTG	3337
1401	ralaThrValSerThrGluGlnProAlaLysGluThrSerSerAsnValG	1418
3338	CGGACCGCGCC.....	3348
1418	luGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrThrMet	1434
3348	3348

1435 ThrGlnThrAlaGluLysSerIspSyrProGlnMetGlnThrValThcGl 1435
1436CGGCAC.....GCAGCGGGGAAATTCGC 3373
1437 ||||| ||| : : : : : |||
1438 uaaNsrPargGlnProGluAlaSnThrValAlaAspAsnSerValAla 1468
1439 GCATTTCGACGGCGGAGGAAAGAAAAACGG.....GTGCAG 34111
1440 : : : : : ||||| : : : : : ||| : : : : :
1441 snAsnSerGlnSerSerGlnSerLysSerArgArgArgSerValSer 1484
1442 : : : : : ||||| : : : : : ||| : : : : :
1443 GCGGATTAAGACACGCCCTTGGCGTAAACAGCGCGAGGAAACCCGCGC 34616
1444 ||| : : : : : ||| : : : : : ||| : : : : :
1445 GlnProLysGlnThrSerAlaGluGlnThrThrValAlaSerThcGlnGl 1501
1446 GCGTACC.....ACCGCTTCCCGCGCGCGCGCG 34939
1447 : : : : : ||| : : : : : ||| : : : : :
1448 uTrpThrValAspAsnSerValSerThrProLysProLysArgSerArgTrp 1518
1449 : : : : : ||||| : : : : : ||| : : : : :
1450 CCGCGCGGGATTTCGGCGCAACCGGACCCGACCCGCAACCCGAC 35438
1451 : : : : : ||||| : : : : : ||| : : : : :
1452 TrpArgSerValGlnThrAsnSerTyGlnProValGlnLeuProThr 1534
1453 3543
1454 GlnAsnAlaGluAsnAlaGlnAsnValGlnSerGlnAsnAsnValAlaAs 1551
1455CGGACCTGATCAGCCGTTATGCCATACCGGTT 35777
1456 nSerGlnProAlaLeuAlaGlnLeuThrSerTyLysSnThrAlaValAl 1568
1457 : : : : : ||||| : : : : : ||| : : : : :
1458 TCAGTAATTTTCGCG.....ACCGTACACGCGTTTC 36121
1459 : : : : : ||| : : : : : ||| : : : : :
1460 LeSerSnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlnLys 1584
1461 : : : : : ||| : : : : : ||| : : : : :
1462 GCGCTACAGAGCAATTCGACGCGCTG.....TTTGGCAAGACCGCGCA 36539
1463 : : : : : ||||| : : : : : ||| : : : : :
1464 AlaValSerGlnHisLeSerGlnLeuGlnLysLysAsnGlnGlnTy 1601
1465 : : : : : ||||| : : : : : ||| : : : : :
1466 TrpValThrLeuSerAsnThrSerMetAsnThrLysSnTySerGlnGluG 1618
1467 : : : : : ||||| : : : : : ||| : : : : :
1468 ATTTCGCGCCCTACGCGCAACAACCGCATCGGCAATCGATCGAG 37599
1469 : : : : : ||| : : : : : ||| : : : : :
1470 TrpTyArgTrpPheSerLysSerThrGlnThrGlnLeuGlnTyArgP 1634
1471 : : : : : ||||| : : : : : ||| : : : : :
1472 AAAACCTCGACGCGG...CGGCTCGGATCTGTTCGCAACACG 38066
1473 : : : : : ||||| : : : : : ||| : : : : :
1474 GlnThrLeuSerAsnAsnValGlnLeuGlnTyAlaPheThrTyValArg 1651
1475 : : : : : ||||| : : : : : ||| : : : : :
1476 GACGGAAACACCTTCGACGAGCGGATCGGCAACTCGGACGCGCTGCC 38536
1477 : : : : : ||||| : : : : : ||| : : : : :
1478 GlnSerAsnAsnPheAspLysAlaSerSerLysSnThrLeuAlaGln 1668
1479 : : : : : ||||| : : : : : ||| : : : : :
1480 ACGGTCCGCTTTCGGCGATTCGCGCATCGGACGCTTCGACATCGGCATC 39060
1481 : : : : : ||||| : : : : : ||| : : : : :
1482 AlaSnThrGlnSerLysTyGlnAlaAspAsnHisThrTyLeuGlnLys 1684
1483 : : : : : ||||| : : : : : ||| : : : : :
1484 AGCGCGCGCGCGGTTTATGTAAGCGCAACCTTCGACGCGCATCGACAG 39566
1485 : : : : : ||||| : : : : : ||| : : : : :
1486 AspLeuGlnTyGlnLysPheGlnSerAsnLeuGlnThrAsnAsnAl 1701
1487 : : : : : ||||| : : : : : ||| : : : : :
1488 GAAATTCGCGCGCGGCTCGCATCGGATCGGACGATCGCAAGATACGCG 4006
1489 : : : : : ||||| : : : : : ||| : : : : :
1490 AlaSerPheAlaAlaGlnHisThrAlaGlnThrLeuThrAlaGlnLysAla 1718
1491 : : : : : ||||| : : : : : ||| : : : : :
1492 CAGGTTTCGCGGATTCGCGCATCGAAGCGCATCGGACGCGCAACGCGAT 4056
1493 : : : : : ||||| : : : : : ||| : : : : :
1494 HeuSnLeuGlnPheAlaValLysProThrPheValGlnTyLysSer 1734
1495 : : : : : ||||| : : : : : ||| : : : : :
1496 TTCGTCGAAAAAGCGATTACCGATACGAAAAAGCGAATATCGCACCC 4106
1497 : : : : : ||||| : : : : : ||| : : : : :

```

1735 TyrLeuSerAsnAlaaspheAlaLeuAlaGlnaspArgIleLysValas 1751
4107 GGGCTTCATTCACACCGCTACCGGGCGCATTAAGGCAATATTCAT 4136
1751 nProIleSerValLysThrAlaPheAlaGlnValaspLeuSerYrThr 1768
4157 TCAAACCGCGCCACACATTCATCCATTCATTTGAGCGCTCCAT 4206
1768 YrHISLeuGlyL...PheSerIleThrProIleLeuSerAlaArgYr 1783
4207 ACCGATGCCGCTCCGCGCAAGTCGAGCGCGCTCAATACCGCGAT 4256
1784 ..AspAlaasnGlnGlyAsnGlyLysIleasnValSerValTyrasp 1799
4257 GCGCGAGATTCGGCAACACCGCGAGTGGGAAATGGCGCGTAAACGCG 4306
1799 eAlaIyrAsnValGluasnGlnGlnGlnIyrAsnAlaGlyLeuLys 1816
4307 AAATCAAGGTTTCACGCTGCTCCACGCTGCCGCGCGCAAGGCGCG 4356
1816 ySTyriHISnValLysLeuSerLeuIleGlyLysLeuThrLysAlaLys 1832
4357 CAATTGGACGCGACACGCGCGCGCATCAATTAGCGCTAC 4398
1833 GlnAlaGlnLysGlnLysThrAlaGlnValLysLeuSerPhe 1846
seq_name: /cgn2_6/prodata/1/aa/PCTUS_COMB.pep:PCT-US95-10661A-6

```

```

seq_documentation_block:
? Sequence 6, Application PC/US9510661A
? GENERAL INFORMATION:
? APPLICANT: Washington University, et al.
? TITLE OF INVENTION: Haemophilus Adherence and Penetration Protein
? NUMBER OF SEQUENCES: 9
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Flehm, Honbach, Test, Albritton & Herbert
? STREET: 4 Embarcadero Center, Suite 3400
? CITY: San Francisco
? STATE: California
? COUNTRY: United States
? ZIP: 94111-4187
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: PCT/US95/10661A
? FILING DATE: 16-Aug-1995
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: US 08/296,791
? FILING DATE: 25-Aug-1994
? CLASSIFICATION:
? ATTORNEY/AGENT INFORMATION:
? NAME: Treacartin, Richard F.
? REGISTRATION NUMBER: 31,801
? REFERENCE/DOCKET NUMBER: FP-59941/RF
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (415) 781-1989
? TELEFAX: (415) 398-3249
? TELEX: 910 277299
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1848 amino acids
? TYPE: amino acid
? TOPOLOGY: unknown
PCT-US95-10661A-6

```

```

alignment_scores:
Quality: 1513.00      Length: 1916
Ratio: 1.519          Gaps: 60

```

Percent Similarity: 51.983 Percent Identity: 25.418

alignment_block:

US-09-303-518d-653 x PCT-US95-10661A-6 ..

Align seg 1/1 to: PCT-US95-10661A-6 from: 1 to: 1848

```

64 CGGTCGCGCGCGCTTACCATATGCGTGGTGGCATTCGCG 113
5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThr 21
114 CCAAGCGCGCGCGGACACATTATTCGGCATCAATCAATATATC 163
21 oYrThrGlnAlaAlaLeuValArgaspPalaSerYrGlnIlePhe 38
164 GCGACTTTCGCGCAAAATAAGGCAAGTTCGAGTGGCGCGCAAGAT 213
38 rGaspheAlaGlnLysnLysGlyLysPheSerValGlyAlaThrAsnVal 54
214 GAGGTTTACACAAATAAGGAGGAGTGGTGGCAATTCATGACGAAC 263
55 GluValArgaspLysLysasnGlnSerLeuGlySerAlaLeuProAsn 71
264 C...CGATGATGATTTTGTGTGTATG...CGTAAGCGGTGGCG 307
71 YrLeuPheIleAspPheSerValAlaAspValaLysArgIleAla 88
308 CATTGGCGCGCGCATCAATATTCGAGCGTGGCAT...AAGCGCGCG 354
88 hrLeuValAsnProGlnTyrValValGlyValLysIleValSerAsnGly 104
355 TATACAAATGTTGATTTTGTGT...GCGAGGAGGAAGCA 389
105 ValSerGlnLeuHisPheGlyAsnLeuAsnGlyAsnMetAlaAsnGly 121
390 TCCCGATCAGCACCGC...TTTCTTCAAAATG 421
121 nAlaLysSerHisArgaspValSerSerGlnLysnArgYrYrThrV 138
422 TGAAGAATAATATAT...AAGCA 444
138 AluGlnLysAsnAsnPheProThrGlnAsnValThrSerPheThrLysGlu 154
445 GGGACTAAGCGCGCATTCATGCGCGATTCATATCCGCGTTTGA 494
155 GluGlnAspAlaGlnLysArgArgLysAspYrYrMetProArgLeu 171
495 CAATTGTTCACAGATGACAGACCTGTGAG...ATGACCATTTATATG 541
171 pLysPheValThrGluValAlaProIleGlnLysSerThrAlaAsnAsn 188
542 ATGGGTGGAATACGCTGATTAATATACCGTGGTGTGGAATC 591
188 snLysGlnGlyLysAsnAsnSerAspLysYrProAlaPheValArgLeu 204
592 GGAGCAGCGACACATAT...TG 611
205 GlySerGlyThrGlnPheIleTyrLysLysGlySerArgYrGlnLeu 221
612 GCGGTCTGATGAAGAGAACCAATACCGCGAAATGATATCAT... 657
221 eLeuThrGlnLysAspLysGlnLysAsnLeuArgAsnTrpAspValG 238
658 .....ATTGACGCGCATTCGCTGGCTGCGT 687
238 LysLysAsnLeuGlnLeuValGlyAsnAlaTyrThrYrGlyLeuAla 254
688 GGCATATCCTTTGCACAAATGATCAGTGGTGGCAGTCACTATGAG 737
255 GlyThrProTyrLysValAsnHisGlnAsnGlyLeuIleGlyPheG 271
738 TAGCGAAAAATTAACATAGC...CCATATGTTTATTACACACAGA 783

```



```
271 yAsnSerIysGluGlnHisSerAspProIysGlyIleuSerGlnAsp 288
784 .....GGCTCATTTGGCGACAGTGGCTGCACCAATGTTATC 819
288 roleuThrAsnTyrAlaValIleuGlnAspSerGlySerProIeuPheVal 304
820 TATGATGCCCAAAAGCAAAAGTGTATTAATGAGGTATTGCAACAGG 869
305 TyrAspArgGlnIysGlyIysTrieuPheIeuGlySerTyrAspPheTr 321
870 CAACCCCTATAGAAAACCAATCGCTTCACCTAGTTCGTAAAGATT 919
321 PALGilyTyr.....AsnIysIysSer 329
920 GCTTCTATGATAATCTTCTGAGATACCATTCAGTATCTAGCA 969
329 rPgIngluTyrAsnIleTyrIysHisGluPheAlaGluIysIleTyrGln 345
970 CCACATCAAAATGGAAATACTTTTAAACGACAATAAATGGCGCAGG 1019
346 GlnTyr.....SerAlaGln 350
1020 AAAATGCGATGCCAAACATAACACTATCTCTACCTTATAGATTAAAA 1069
350 ySerIeuIleIysSerAsnThrGlnTyrThrPglAlaIleThrGlySer 367
1070 CACGACCGCTT.....CAATGTTTAATGTTCTTATTCGAG 1107
367 hrSerThrIleThrGlyGlyIgluIurProIeuSerValAspIeuThrAsp 383
1108 ACACGACAGAACCTGTTTATCATGCTGCAGAGTGGGTCAACACTTATCG 1157
384 GlysAspIysPro..... 388
1158 ACCGACAGCTGATATATGAGAAATATTCTTATTGACAAAGAAAG 1207
389 .....AsnHisGlyIysSerIleThrIeu.....LysGlySerG 400
1208 GTGATTTGATACTTACGACGACATCAACCAAGCGCGGCGTGTAT 1257
400 LThrIeuThrIeuAsnAsnHisIleAspGlnGlyAlaGlyIleuPhe 416
1258 TTTGAGGTAATTTTACGGTC...TCGCTAAAAAACAAGAAACGTGCA 1304
417 PheGluIysAspTyrGluValIysGlyThrSerAspSerThrThrTyr 433
1305 AGCGCGGCGCTTCATATCACTGATGGCAGTACCGTACTTGGAAAGTAA 1354
433 sGlyAlaGlyValSerValAlaAspGlyIysThrValThrTrieuValH 450
1355 ACGGCGTGGCAACGACCGCTGCCAAATCGGCAAGGACGCTGCTG 1404
450 IsAsnProIysTyrAspArgIeuAlaIysIleGlyIysGlyThrIeuVal 466
1405 GTTCACGCCAAAGGGAACCAAGGCTCGTACGCTGGCGACGCTAA 1454
467 ValGluIysGlyIysAsnGluIleuIeuIysValGlyAspGlyTh 483
1455 AGTATTTGATGACGACGCGGACGATCAAGCAAAAAACAACCTTTA 1504
483 rValIleuIysGlnIysAlaAspAlaAsnAsnIysValGlnAlaPheS 500
1505 GTGAATCGCTTGTGACGCGGACGAGGAGGCAACTGATGCGCAT 1554
500 erGlnValIleValSerGlyIysSerThrIeuValIleuAsnAspAsp 516
1555 AATCATGTTCAACCCGACAAACTTATTTGGCTTTCGCGCGACGTTT 1604
517 LysGlnValAspProAsnSerIleTyrPheGlyPheArgGlyGlyArgLe 533
1605 GGATTTGAAGGGCATTCGCTTTCGCTTCACCCGATTCAAATAACGATG 1654
533 uAspIeuAsnGlyAsnSerIeuThrPheAspHisIleArgAsnIleAspA 550
1655 AAGGGGATGATTGTCAACCAATCAAGACAAGAAAGATCCACCGTTACC 1704
550 sPglAlaAlaGlyValIleAsnHisAsnMetThrAsnThrSerAsnIleThr 566
1705 ATTACAGCGAATTAAGATATTTACTACA.....ACCGCAATTA 1742
567 IleThrGlyIuSerIeuIleThrAsnProAsnThrIleThrSerTyrAs 583
1743 C..... 1743
583 nIleGluAlaGlnAspAspAspHisProIeuArgIleArgSerIleProT 600
1743 ..... 1743
600 yThrGlnIeuIyrPheAsnGlnAspAsnArgSerTyrTyrThrIeuIys 616
1743 ..... 1743
617 LysGlyAlaSerThrArgSerGluIeuProGlnAsnSerGlyGluSerAs 633
1743 ..... 1743
633 nGluAsnTrieuIyrMetGlyArgThrSerAspAlaAlaIysArgAsnV 650
1744 .....AACAACTTGATAGCAAAAAAGAAATGGCTACACAGGTGTTT 1788
650 alMetAsnHisIleAsnAsnGluArgMetAsnGlyPheAsnGlyTyrPhe 666
1789 GGGGAGAAAGATGCAACAAACG...AACGGGCGGCTCAATCTGAATTA 1835
667 GylGluGluGluThrIysAlaThrGlnAsnGlyIysIleuAsnValThrPh 683
1836 CCACCGGAAAGAACCGGATCGCACTTACTGCTTCGCGGCAACAAT 1885
683 eaSnGlyIysSerAspGlnAsnArgPheIeuIeuThrGlyIThrAsnL 700
1886 TAAACGCAATATCAGTCAACAAACGCAAACTGTTTTCGCGGCGAGA 1935
700 euAsnGlyIysIleuAsnValIleGlyIysGlyThrIeuPheIeuSerGlyArg 716
1936 CCGACACCGCACCGCTACATCATTTAGAGCGGGGTGCAAAATGGA 1985
717 ProThrProHisAlaArgAspIleAlaGlyIleSerSerThrIyIysAs 733
1986 ArgT.....ATCCCAAGAGAAATCGTGGGACAAACGATTGATCG 2029
733 pProHisPheThrGluAsnAsnGluValAlaValGluAspAspTyrIleA 750
2030 ACCGACATTTAAAGCGGAAACCTTCATATTCAGGCGGACAAACGGTG 2079
750 snArgAsnPheIysAlaIleThrThrMetAsnValThrGlyAsnAlaSerIeu 766
2080 GTTTC...CGCAATGTTGCCAAAGTGAAGGCGATTTGCAATTAAGCA 2126
767 TyrSerGlyArgAsnValAlaAsnIleThrIserAsnIleThrAlaSerAs 783
2127 TCACGCCCAAGCAATTTTGGTGGTGGCACCGCATTAAGCCACACAACT 2176
783 nAsnAlaGlnValHisIleGlyTyr.....LysThrGlyAspThrValC 798
2177 GTACAGCTTCGAGCTGAGCGGCTGACAACTGTTCACGAAAAAAACAT 2226
798 yValAlaArgSerAspTyrThrGlyTyrValThrCysHisAsnSerIeu 814
2227 ACCGACGATTAAGTATGCTTCATTTGACGACGACGACGACGACG 2276
815 Ser...GluIysAlaIleuAsnSerPheAsnProThrIleAsnGlyAs 830
2277 TGTCAAGCTTCGCGATCACGCTCATTTAAATCTCACAGACTTGCACAC 2326
830 nValAsnIeuThrGluAsnAlaSerPheThrIeu...GlyIysAlaAsnL 846
```


1418 IuGlnProAlaProGluAsnSerIleAsnThrGlySerAlaThrThMet 1434
3348 3348
1435 ThrGlnThrAlaGluLysSerAspLysProGlnMetGluThrValThrGln 1451
3349CGCGAG.....GCAGCGGGGGAATAATCCG 3373
1451 uAsnAspArgGlnProGluAlaAsnThrValAlaAspAsnSerValAla 1468
3374 GCATTATCGAGCGGAGAGAGAGAAAAACGG.....GTCCAG 3411
1468 sAsnAsnSerGluSerIleSerIleSerIleSerIleSerIleSerIleSer 1484
3412 GCGGATTAAGACACCGCTTGGCGAAACAGCGGAGCGGAGAAACCGCGCC 3461
1485 GlnProLysGluThrSerAlaGluGlnThrThrValAlaSerThrGlnGln 1501
3462 GCGTACC.....ACCGCTTCCCGCGCGCGCGCGCGCG 3493
1501 uThrThrValAspAsnSerValSerThrProLysProArgSerArgArgT 1518
3494 CCGCGCGGAGATTGCCGACCGGACCGCCGACCGGACCGGACCGGACCGCAG 3543
1518 hrArgArgSerValGlnThrAsnSerTyrGluProValGluLeuProThr 1534
3543 3543
1535 GluAsnAlaGluAsnAlaGluAsnValGlnSerIleLysAsnValAlaAs 1551
3544CGCGACTGATCAGCGCTTATCGCAATAGCGGTT 3577
1551 nSerGlnProAlaLeuArgAsnLeuThrSerLysAsnThrAsnAlaValI 1568
3578 TGAGTGAATTTCCGCC.....ACGTCACACAGCGTTTC 3612
1568 IeSerAsnAlaMetAlaLysAlaGlnPheValAlaLeuAsnValGlyLys 1584
3613 GCCGTACAGAGCAATTGACCGCGTG...TTGCCGAGACGCGCGCA 3659
1585 AlAlaLysGlnHisIleSerGlnLeuGlnMetAsnGlnGlyGlnTyr 1601
3660 CCGCGTTTGCAACAGCGGATCGGACACCAACTACCGTTTCGCAAG 3709
1601 rAsnValTyrIleSerAsnThrSerMetAsnLysAsnTyrSerSerGln 1618
3710 ATTCCGCGGCTACCGCCCAACAAACGACTCGCCCAATCGGATACAG 3759
1618 IuThrArgArgPheSerSerLysSerThrGlnThrGlnLeuGlyTyrPasp 1634
3760 AAAAACCTCGAGCGGG...CGGTCGCGACTCTGTTTCGACAGACCG 3806
1635 GlnThrIleSerAsnAsnValGlnLeuGlyValPheThrTyrValAla 1651
3807 GACCGGAAACCTTCGACAGCGGATCGCAACTCGGACGCGCTTCGCC 3856
1651 gAsnSerAsnAsnPheAspLysAlaSerSerLysAsnThrLeuAlaGln 1668
3857 ACGGTGCGGCTTTCGGCATACGCGGATCGGATCGGATCGGATCGGAT 3906
1668 aLanPheTyrSerLysTyrTyrAlaAspAsnHisTyrPheGlyTyrIle 1684
3907 AGCGCGGCGCGGCTTTCAGCGGAGCGCTTCAGACGCGCATACGAGG 3956
1685 AspLeuGlyTyrGlyLysPheGlnSerAsnLeuGlnThrAsnAsnAla 1701
3957 CAATAACCGCGCGCGCTGCTGCATTACGCGATTCAGGCAAGTACCGCG 4006
1701 aLysPheAlaArgHisThrAlaGlnIleGlyLeuThrAlaGlyLysAla 1718
4007 CAGGTTTCGCGGATTCGCGATCGCAACGCGATCGGACGCGCGTAT 4056
1718 hAsnLeuGlyAsnPheAlaValLysProThrValGlyValArgTyrSer 1734

4057 TTCGTCAAAAAGCGGATTACGATACGAAGAAACGTCATATCGCCACCC 4106
1735 TyrLeuSerAsnAlaAspPheAlaLeuAlaGlnAspArgIleLysValAs 1751
4107 GGGCGCTTGCAATTCACCGCTACCGCGCGGCGCATTAAGCGAGATTATCAT 4156
1751 nProIleSerValLysThrAlaPheAlaGlnValAspLeuSerTyrThrT 1768
4157 TCAAAACCGCGCACACATTTCCATTCAGCGCTTATTGACGCTGCTAT 4206
1768 yTrHisLeuGlyGlu...PheSerIleThrProIleLeuSerAlaArgTyr 1783
4207 ACCGATCCGCTTCGCGGAAAGTCGCAAGCGGCGTCAATACGCGCTAT 4256
1784 ...AspAlaAsnGlnLysnGlyLysIleAsnValSerValTyrAspH 1799
4257 GCGCGAGATTTCGCAAAACCGCGAGTCGGAATGGGCGTAAACGCGCG 4306
1799 eAlaTyrAsnValGluAsnGlnGlnGlnIleTyrAsnAlaGlyLeuLysLeu 1816
4307 AAATCAAAGTTTCACGCTGCTCCCTCCACGCTGCGCGCGCAAGGCGCG 4356
1816 yStrHisAsnValLysLeuSerLeuIleGlyGlyLeuThrLysAlaLys 1832
4357 CAATTGGAAGCGGACGACGCGGCGCATCAATTAAGCTAC 4398
1833 GlnAlaGluLysGlnThrAlaGluValLysLeuSerPhe 1846
seq_name: /cgn2_6/plodata/1/lae/5B.COMB.pep:US-08-728-470-9
seq_documentation_block:
; Sequence 9, Application US/08728470
; Patent No. 5928651
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF INVENTIONS: of No. 5928651-Typeable Haemophilus
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728,470
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Belkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-633
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:


```

1334 GTACCGTTACTTGAAAGTAAACGGC..... 1359
      ::::::::::::::::::::
489 LysValAsnPhelYsLeuAsnAlaSerSerAsnIleGlnThrPro 505
      ::::::::::::::::::::
1360 .....GTGGCAACGACCGCTGTCCAAATCGCAAAAGCAGCGCT 1400
      ::::::::::::::::::::
506 GlValIleIleYsSerGlnAsnPhenAsnValSerGlySerThrIle 522
      ::::::::::::::::::::
1401 GCTGCTCAGCCAAAGG.....GAAACC 1426
      ::::::::::::::::::::
522 uAsnLeuYsAlaGluGlySerThrGlnThrAlaPheserIleGlnAsn 539
      ::::::::::::::::::::
1427 AAGCTCGGTAGCGTGGCGAGGTAAAGTCATCTTAGATCAG...CAG 1473
      ::::::::::::::::::::
539 sPleuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGlnValGlu 555
      ::::::::::::::::::::
1474 GCGGACGATCAA..... 1485
      ::::::::::::::::::::
556 GlyThrAspSerArgValAsnGlyValAlaAlaYsLysAsnIleTh 572
      ::::::::::::::::::::
1486 .....GGCAAAACAAAGCCTTTAGTGAAA 1510
      ::::::::::::::::::::
572 rPhelYsGlyGlyAsnIleThrPhelGlySerGlnLysAlaThrGln 589
      ::::::::::::::::::::
1511 TCGGCTTGTCAGCGGCGAGCGGTGCACTGAATGCCGATATCAG 1560
      ::::::::::::::::::::
589 Le.....LysGlyAsnValThrIleAsnLysAsnThrAsn 600
      ::::::::::::::::::::
1561 TTCAACCCCGCAAACTCTATTGTGGCTTGGCGGAGCGTGTGGATT 1610
      ::::::::::::::::::::
601 AlaThr.....LeuArgIlyAlaAsnPhenAlaG 610
      ::::::::::::::::::::
1611 GAACGGGCGATTGGCTTGCCTCCACCGCATTCAAATACGATGAAGGG 1660
      ::::::::::::::::::::
610 uAsnLysSerProLeuAsnIle.....AlaGly 620
      ::::::::::::::::::::
1661 CGAGTATTGTCAACCAACAT...CAAGCAAAAGATCCACCGTTACATT 1707
      ::::::::::::::::::::
620 snValIleAsnAsnGlyAsnLeuThrThrAlaGlySerIleIleAsnIle 636
      ::::::::::::::::::::
1708 ACAGCGCATTAAGATATTACTACAAACCGCAAT..... 1740
      ::::::::::::::::::::
637 AlaGlyAsnLeuThrValSerLysGlyAlaAsnLeuGlnAlaIleThrAs 653
      ::::::::::::::::::::
1741 .....AACCAACCTTGATAGCAAAA 1762
      ::::::::::::::::::::
653 nTyThrPhenAsnValAlaGlySerPhenAspAsnAsnGlyAlaSerAsnI 670
      ::::::::::::::::::::
1763 AAGAATTGCTACACAGGTTGTTGGCGAGAAAGATGCAACCAAAAGC 1812
      ::::::::::::::::::::
670 leSerIleAla...ArgIlyGlyAlaLysPhelYsAspIleAsnAsnThr 685
      ::::::::::::::::::::
1813 AACGGGGGCTCAATCTGAATTACCAACCGAAGAACGGATCGCACTTT 1862
      ::::::::::::::::::::
686 SerSer...LeuAsnIleThrThrAsnSerAspThrThrIleArgThrI 701
      ::::::::::::::::::::
1863 ACTGCTTCCGGCGGAGAACAAATTAAACGCGCATATACGCAAAACAAG 1912
      ::::::::::::::::::::
701 e.....IleYsGlyAsnIleSerAsnLysSerG 711
      ::::::::::::::::::::
1913 GCAAACTGTTTTCAGCGGAGACCGACCGCGCTTACATTCATTTA 1962
      ::::::::::::::::::::
711 LysPleuAsnIleIleAspLysSer...AspAlaGlnIleGlnIle 726
      ::::::::::::::::::::
1963 GGAAGCGGTTGTCAAAATGAAGGTATCCACAAAGAGAAATCGTGTG 2012
      ::::::::::::::::::::
727 GlyLysAsnIleSerGlnLysGlu.....GlyAsnLeuThrI 739
      ::::::::::::::::::::
2013 GGACAAGGATTTGATGAC.....CGCATTTTAAAGCGGAAA 2050
      ::::::::::::::::::::
739 eSerSerAspLysValAsnIleThrAsnGlnIleThrIleLysAla.... 754
      ::::::::::::::::::::
2051 ACTTCATATTCAAGGGCGGACAAGCG..... 2076
      ::::::::::::::::::::

```

```

755 .....GlyValIleGlyGlyArgSerAspSerSerGlnAlaGluAsnAla 769
      ::::::::::::::::::::
2077 .....GTGGTTCCCGCAATGTTGCCAAAGTGAAGCGGATTTGCATT 2120
      ::::::::::::::::::::
770 AsnLeuThrIleGlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnI 786
      ::::::::::::::::::::
2121 AAGCAATCAGCCCGCAAGCACTTTGCGTGCAGCCGATCAAGCAACA 2170
      ::::::::::::::::::::
786 eSerGlyPhenAsnLysAlaGlnIleThrAlaLysAsnGlySerAspLeu 803
      ::::::::::::::::::::
2171 CAATCTCAGCACTGCGAGCTGAGCGGTCTGACAGTTGTACCGAAAA 2220
      ::::::::::::::::::::
803 hrIleGlyAsnAlaSerGly.....GlyAsnAlaAspAlaLys 815
      ::::::::::::::::::::
2221 ACCATTACGAGCATAAAGTATGCTTCATTGAGCAAGACCGCATCAG 2270
      ::::::::::::::::::::
816 LysValThrPhenAspLysValLysAspSerLysIleSerThrAsp... 831
      ::::::::::::::::::::
2271 AGGCATGTCAGCCTTCCGATCAGCCTCATTTAAATCTCACAGACTTG 2320
      ::::::::::::::::::::
831 YHisAsnValThrLeuAsnSerGlu.....ValL 841
      ::::::::::::::::::::
2321 CCACACTCAACGCGCAATCTTAGTGCAGCGGAGAC.....ACGCACTAT 2364
      ::::::::::::::::::::
841 YThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerThrGlyLeu 857
      ::::::::::::::::::::
2365 ACGGTTACG...CGCAACGCCACCCAAAAGCGCACTCAGCCTGCTGG 2411
      ::::::::::::::::::::
858 ThrIleSerAlaLysAspValThrValAsnAsnAsnValThr..... 871
      ::::::::::::::::::::
2412 CAATGCCCAAGCAACATTTAATCAAGCCACATTTAAAGCGCAACACATG 2461
      ::::::::::::::::::::
872 ...SerHisLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThrT 887
      ::::::::::::::::::::
2462 CTTGGACAAATGCTTCATTAACTAAGCAACAGCCGTA..... 2502
      ::::::::::::::::::::
887 hrLysGlnGlyThrThrIleAsnAlaThrThrGlySerValGluValThr 903
      ::::::::::::::::::::
2503 .....CAAAACGCGACT 2516
      ::::::::::::::::::::
904 AlaGlnAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrYa 920
      ::::::::::::::::::::
2517 GACGCTTTCGACCAAC.....GCTAAGGCAACGTAAGCAATTCGCGAC 2560
      ::::::::::::::::::::
920 LThrAlaThrGluAsnLeuValThrThrLysAsnAlaValIleAsnAlaT 937
      ::::::::::::::::::::
2561 TCACAGCGCAATGTCTCCTAGCCGATAGCGAGATTCATTTTGAAAA 2610
      ::::::::::::::::::::
937 hrSerGlyThrValAsnIleSerThrLys..... 946
      ::::::::::::::::::::
2611 AGCGCTTACCGGAAATCAAGCGGCGGCAAGATACGCA..... 2652
      ::::::::::::::::::::
947 .....ThrGlyAspIleLysGlyIleGluSerThrSerGlyAs 960
      ::::::::::::::::::::
2653 ..TTACACTTAAAGACGGAATGAGACGCTGCGGTGCGGACAGGAATAG 2701
      ::::::::::::::::::::
960 nValAsnIleThrAlaSerGlyAsnThrLeu.....LysValS 973
      ::::::::::::::::::::
2702 GCAATTTAACTTGACAAACGCCACATTACCTCAATTCGCGCTATGCA 2751
      ::::::::::::::::::::
973 eAsnIleThrGlyGlnAspValThrValThrAlaAsp..... 985
      ::::::::::::::::::::
2752 CACGATGCGGACGCGGCAACCGGAGTGGGAGATGCGCGCGCG 2801
      ::::::::::::::::::::
986 .....AlaGlyAlaLeuThrThrAlaGly..... 994
      ::::::::::::::::::::
2802 CCGTTGCGCGCTTCCTATTATCGTTACGCGGCAACTTCGCGCAAT 2851
      ::::::::::::::::::::
995 .....SerThrIleSerAlaThrThrGlyAsnAla..... 1004
      ::::::::::::::::::::
2852 CCGGTTTCAACAGCGTGAACGGTAACGCAATTAACAGGTACGAGAA 2901
      ::::::::::::::::::::

```

```

1005 .....AsnIleThrThrLysThrGlyAspIleasnGlyLys..... 1016
2902 TTCCGGTTTATGTCGAACTCTTCGGCTACCGACGGCAAAATTGAAGCT 2951
1016 ..... 1016
2952 GGGGGAAGTTCCGAAAGCCTTAC.....ACCT 2980
1017 ValGluSerSerSerGlySerValThrIleValAlaThrGlyAlaThrL 1033
2981 TGCGTGTCAACATACCGGCAAGAACCCCTAGTCTCCGACCAATTGACG 3030
1033 eulAlaValAlGlyAsnIleSerGlyAsnThrValThr.....IleThr 1046
3031 GTAGTGAAGAAAGAAAGACACACCGCTCCGAAATCTTAATTTCAC 3080
1047 AlaSpSerGlyLysLeuThrSerThrValGlySerThrIleasnGlyTh 1063
3081 CCTGCAAAAGAACACGTCGATGCCGCGCATGCGCTTATCAGCTTATCC 3130
1063 ..... 1063
3131 GCAAAAGCGGAGTCCGCTGCATATCCGTCGAAAGAACAGAGCTT 3180
1064 .....AsnSerValThrThrSerSerGlnSer 1072
3181 TTCCGCAAACTCGGCAAGGGGCGGAAAGAGCGCGCTTGACGGCAAA 3230
1073 GlysPlleGlyThrIleSerGlyAsnThrValAsnThrAlaThrLase 1089
3331 ACAGCGCAACTT.....GCCGCAAAACACAGCGGCAAAAG 3268
1089 rThrGlyAspLeuThrIleGlyAsnSerAlaLysValGluAla...LysA 1105
3269 ACAACGGCGCAAGCCCTTGACGCGCTGATTCGCGCGCGCGCAATGCCAC 3318
1105 snGlyAlaAlaThrLeuThrAla....GluSerGlyLysLeuThrThr 1119
3319 GAAAGGCGAAGAGTGT.....GC 3338
1120 GlnThrGlySerSerIleThrSerSerAsnGlyGlnThrThrLeuThrAl 1136
3339 CGAACCGCGCGGCGAGCGGCGGAAATGCC..... 3372
1136 aLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnValThrLeuA 1153
3373 .....GCCATTATGACGCGCGGAGGAAAGAAACCGGTCACGCG 3414
1153 snThrThrGlyThrLeuThrThrThrGlyAspSerLys...IleAsnAla 1168
3415 GATTAAGACACCGCTTGCGAAGACGCGGCAACCGGCGCGGCG 3464
1169 ThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeuAspGlyAl 1185
3465 TACGACCGCGCTTCCCGCGCGCGCGCGCGCGGAGATTGCGCGCAAC 3514
1185 aAlaSerGly..... 1188
3515 CGCAGCCCAACCGCAACCCCAACCGGCGGACCTGCATCAGCGCTTAT 3554
1189 .....AspArgThrValAlaAsnAlaThr 1196
3565 GCCAATACGGGTTGAGTGAATTTCCGCCACGCTCAACAGCGTTTCGC 3614
1197 AsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSerSerValas 1213
3615 CGTACGAGCAATTTGAC.....CGCGTGTTCGCGAAG 3649
1213 nIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluA 1230
3650 ACCGCGGCAAGCGGCTTGAGAACGCGCATCCGG...GACACCAACAC 3696
1230 snGlyArgAsnThrValArgLeuArgGlyLysGluIleAspValLysTyr 1246

```

```

3697 TACGTTGCAAGATTTCGCGGCTACCGCCCAACACCGCACTGCGCA 3746
1247 Ile.....GlnPro.GlyValAlaIleSerValGluGluValIleGluAlaL 1261
3747 AATCGTATGCAAGAAAC.....TCG 3769
1261 ysArgValLeuGluValLysAspLeuSerAspGluArgGluThr 1277
3770 GCAGCGGCGCGTCGCGCATCTGTTTTCGACACACCGGCGGAAACAC 3819
1278 LeuAlaLysLeuGlyValSerAlaValAlaArgPheValGluProAsnAl 1294
3820 TTGACAGCGCGCATCGGCACTCGGACGCGCTTGCCCGCGGCTTTT 3869
1294 alleThrVal.....AsnThrGlnAsnGluPheThrThrLysProSerS 1309
3870 CCGGCAATACGCGCATCGGCGAGTTTCGACA.....TCGGCATACGCGG 3913
1309 ergLValThrIleSerGluGlyLysAlaCysPheSerSerGlyAsnGly 1325
3914 GCGCGGTTTGTAGCGCGCAGCC 3937
1326 AlaArgValCysThrAsnValAla 1333

seq_name: /cgn2_6/prodata/1/iaa/68_COMB.pep:US-08-719-641-9
seq_documentation_block:
; Sequence 9, Application US/08719641
; Patent No. 6218141
;
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESS: Shoemaker and Mattare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1338 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```


1360GTGGCAACGACCGCGCTGCCAAATCGCAAGGACGCT 1400
::: ::::::::::::::: ||| ::::|
506 GlyValIleIleIysSerGlnAsnPheAsnValSerIylGlySerThrIle 522
1401 GCTGGTTCAAGCCCAAGG.....GAAACC 1426
| ::::|::|::|::|
522 uAsnLeuLysAlaGluSerThrGluThrAlaPheSerIleGluAsnA 539
1427 AAGGCTGGTCAGCGTGGCGAGCGTAAGCATCTAGATCAG...CAG 1473
::: ::::::::::::::: |||:::|
539 sPLeuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGlnValGlu 555
1474 GCGGACGATCA..... 1485
::: |||:::
556 GlyThrSperSerArgValAsnLysGlyValAlaAlaLysLysAsnIleTh 572
1486GGCAAAAAACAACCTTAGTGA 1510
|||:::|::|::|
572 rPheLysGlyGlyAsnIleThrPheGlySerGlnLysAlaThrThrGlu 589
1511 TCGGCTGGTCAGCGGAGCGGCGGTGCACATGATGCCAATACAG 1560
:::|::|::|::|::|::|
589 Le.....LysGlyAsnValThrIleAsnLysAsnThrAsn 600
1561 TTCACCCCGACAAACTCTATTTCGCTTCGCGGCGAGCTTTGATTT 1610
::: ::::|
601 AlaThr.....LeuArgGlyAlaAsnPheAlaGlu 610
1611 GAACGGCATTCGCTTCGCTTCACCCCATCAAAATACCGATGAGG 1660
||| ::::|
610 uAsnLysSerProLeuAsnIle.....AlaGlyA 620
1661 CGATGATTCACCAACCAAT...CAAGACAAGATCCACGTTACAT 1707
:::|::|::|::|
620 snValIleAsnAsnGlyAsnLeuThrThrAlaGlySerIleIleAsnIle 636
1708 ACAGGCATAAAGATATTACTACACCGCAT..... 1740
:::|
637 AlaGlyAsnLeuThrValSerLysGlyAlaAsnLeuGlnAlaIleThrAs 653
1741AACCAACTTGGATGCAAAA 1762
:::|
653 nTyrThrPheAsnValAlaGlySerPheAspAsnAsnGlyAlaSerAsnI 670
1763 AAGAATTCCTACACGCTTGGTGGCGAAGATGCAACCAACAG 1812
:::|::|::|::|
670 IeSerIleAla...ArgGlyGlyAlaLysPheLysAspIleAsnThr 685
1813 AACGGCGGCTCAATCTGAATTACCAACGGAAGACGATCGACTTT 1862
:::|::|::|::|
686 SerSer...LeuAsnIleThrThrAsnSerAspThrThrArgThrIle 701
1863 ACTGCTTCGCGCGGAACAATTTAAACGGCAATATCACGCAAAACAG 1912
: ::::|
701 e.....IleLysGlyAsnIleSerAsnLysSerG 711
1913 GCAACGTGTTTTCAGCGGACGCGACGCGACGCTACATCATTTA 1962
||| ::::|
711 LysPheLeuAsnIleIleAspLysSer...AspAlaGluIleGlnIle 726
1963 GGAAGCGGCTGCACAAATGGAAGTATCCACACGAGCAAAATCGTGTG 2012
|||:::|
727 GlyGlyAsnIleSerGlnLysGlu.....GlyAsnLeuThrIle 739
2013 GGACAGCATTCGATCGAC.....CGCACATTTAAACGGCAAA 2050
:::|
739 eSerSerAspLysValAsnIleThrAsnGlnIleThrIleLysAla.... 754
2051 ACTTCATATTCAGGCGGACACAGG..... 2076
:::|
755GlyValGluGlyLysArgSerAspSerSerGluAlaGluAsnAla 769
2077GTGTTTCCGCAATGTTGCCAAAGTGGAAGCGATTCGATTT 2120

770 AsnLeuThrIleGlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnIle 786
:::|
2121 AAGCAATCACGCCCAAGCAGATTTTGGGTGCGCACCAGCATCAAAACCACA 2170
:::|
786 eSerGlyPheAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeuT 803
2171 CATCTGTACAGCTTCGACCTGAGCGGCGTGCACAGTGTACCGAATAA 2220
|||:::|
803 hIleGlyAsnAlaSerGly.....GlyAsnAlaAspAlaLys 815
2221 ACCATTACCGCAGCAATTAAGTGTATTCATTAGACAAAGCCAGCATCAG 2270
:::|
816 LysValThrPheAspLysValLysAspSerLysIleSerThrAsp...GI 831
2271 AGGCAATGTCAGCTTCGCGATCAGCTTCATTTAATCTCAGAGACTTG 2320
|||
831 yHisAsnValThrLeuAsnSerGlu.....ValL 841
2321 CCACACTCAACGCAATCTTAGTCAGCGGAGAC.....ACGCATAT 2364
|||
841 yThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerThrGlyLeu 857
2365 ACGGTTACG...CGCAACGCCACCCAAAGGCAACTCAGCTTGCGG 2411
|||
858 ThrIleSerAlaLysAspValThrValAsnAsnValThr..... 871
2412 CAATGCCCAAGACATTTAATCAAGCCACATTAACGGCAACATACGG 2461
:::|
872 ...SerHisLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThrT 887
2462 CTTCGACATGCTTCATTTAATCTAAGCAACACCGCTA..... 2502
:::|
887 hLysGluLysThrThrIleAsnAlaIleThrGlySerValGluValThr 903
2503CAAACGCACT 2516
904 AlaGlnAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrVa 920
2517 GACGCTTCGCGAAC.....GCTAAGGCAACGTAGCCATTCGCGAC 2560
|||
920 ThrThrAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAlaT 937
2561 TCAACGCAATGTCCTCCCTAGCGCATTAAGGCAATTCATTTTGAAC 2610
:::|
937 hSerGlyThrValAsnIleSerThrLys..... 946
2611 AGCGCTTACCGAATAATCAAGCGGCGCAAGATPACGGCA..... 2652
|||
947ThrGlyAspIleLysGlyIleGluSerThrSerGlyAs 960
2653 TTACACTTAAGAACAGCATGACGCTGCGCGTGGGACAGGATTAAG 2701
|||
960 nValAsnIleThrAlaSerGlyAsnThrLeu.....LysValS 973
2702 GCAATTTAAACCTTGACAACGCCACCATTTACACTAATTCGCTATGCA 2751
:::|
973 eAsnIleThrGlyGlnAspValThrValThrAlaAsp..... 985
2752 CACGATGCGGCGGCGCAACCGGCACTGGCGGAGATCGCCCGCGCG 2801
|||
986AlaGlyAlaLeuThrThrThrAlaGly..... 994
2802 CGTTGCGCGCTTCCTATTATTCGCTTACGCCGCAACTTCGCGAGAT 2851
|||
995SerThrIleSerAlaThrThrGlyAsnAla..... 1004
2852 CCGGTTTCAACACGCTGACGCTTAACGGCAATTAAGCGTACGGGAACA 2901
|||
1005AsnIleThrThrLysThrGlyAspIleAsnGlyLys..... 1016
2902 TTCGCTTATGTGGAAGCTTCGCGCTACCGCAGCGCAATTAAGCT 2951


```

1016 ..... 1016
2952 GCGCAAGATTCGAGCAGCACTTAC..... ACCT 2980
1017 ValGluSerSerSerSerValThrLeuValAlaThrGlyAlaThrL 1033
2981 TGCGTGTCAACATATCCGGCAGCAACACCGTAACTCGAGCAATTGACG 3030
1033 euAlaValAlGlyAsnIleSerGlyAsnThrValThr..... IleThr 1046
3031 GTACTGAGAGAAAGACACACACCGCTGCGAAATCTTAATTTCAC 3080
1047 AlaSerSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThr 1063
3081 CCTGCAAAAGCAACGCTCATGCGCGCATGCGCTTATCAGTTATCC 3130
1063 r..... 1063
3131 GCAAAGACGGCGAGTTCGCTGCATATCCGGTCAAAGAACAGACTT 3180
1064 ..... AsnSerValThrThrSerSerGlnSer 1072
3181 TCGCAAAACTCGGCAAGGGGAGAAACAGAGCCGCTTGACGGCAAA 3230
1073 GlyAspIleGluGlyThrIleSerGlyAsnThrValAsnValThrAla 1089
3231 ACAGGCACAACTT..... GCGCGCAAAACAGAGCGGAAAG 3268
1089 rThrGlyAspLeuThrIleGlyAsnSerAlaLysValGluAla... LysA 1105
3269 ACAACGGCGCAAGCCTTGACGCGCTGATTCGCGCGGCGCAATGCCACC 3318
1105 snGlyAlaIaThrLeuThrAla..... GluSerGlyLysLeuThrThr 1119
3319 GAAAGCGCAGAAAGTGT..... GC 3338
1120 GlnThrGlySerSerIleThrSerAsnGlyGlnThrThrLeuThrAl 1136
3339 CGAAGCGGCGCGGAGCAGCGGGAATGCC..... 3372
1136 aLysAspSerSerIleAlaGlyAsnIleAsnAlaIaAsnValThrLeuA 1153
3373 ..... GGCATTATGACAGCGCGAGAGAGAAAAACGGGTGACGGCG 3414
1153 snThrThrGlyThrLeuThrThrThrGlyAspSerLys... IleAsnAla 1168
3415 GATAAGACACCGCCTTGCGAAGACCGGAGCGGAGGAAACCGCGCGCGC 3464
1169 ThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeuAspGlyAl 1185
3465 TACACCGCGCTTCCCGCGCGCGCGCGCGCGGATTTGCCCAAC 3514
1185 aAlaSerGly..... 1188
3515 CGAGGCGCCCAACCGCAACCCCAACCGCAGCGGACTGATCAGCGCTTAT 3564
1189 ..... AspArgThrValAlaValAsnAlaThr 1196
3565 GCCAATAGCGGTTTGAGTGAATTTCCGCGACGCTCAACAGCGTTTCCG 3614
1197 AsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSerValAs 1213
3615 CGTACAGCAGCAATTGAC..... CGCGTGTTCGCGAAG 3649
1213 nIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIleSerGluA 1230
3650 ACCGCGCAAGCGCTTGGACAGCGGATCCGG... GACACCAACAC 3696
1230 smGlyArgAsnThrValArgLeuArgGlyLysGluIleAspValLysTyr 1246
3697 TACCGTTCCAGATTTCCGCGCTACCGCAACAAACGACCTGCGGCA 3746
1247 Ile..... GlnPro.GlyValAlaSerValGluGluValIleGluAlaL 1261
3747 ATCGGTATGCAAAAACC.....TCG 3769
1261 ysArgValLeuGluLysValLysAspLeuSerAspGluGluArgGluThr 1277
3770 GCAGCGGCGCGCATCGCTGTTTTCGACACACCGGACGGAACACC 3819
1278 LeuAlaLysLeuGluLysAlaSerAlaValArgPheValGluProAsnAl 1294
3820 TTCGACAGCGCATCGCACTCGGCGACGCTTGCCCGCGGCTTTT 3869
1294 aIleThrVal..... AsnThrGlnAsnGluPheThrThrLysProSer 1309
3870 CGGCGCAATCGGCATCGGCGAGTTCGACA..... TCGCATCAGCGCG 3913
1309 eArgLysValThrIleSerGluGlyLysAlaCysPheSerSerGlyAsnGly 1325
3914 GCGCGGTTTGTAGTCGCGCAGCC 3937
1326 AlaArgValCysThrAsnValAla 1333
seq_name: /cgm2_6/ptodata/1/lae/5B_COMB pep: US-08-617-697-9
seq_documentation_block:
: Sequence 9, Application US/08617697
: Patent No. 5977336
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins
: TITLE OF INVENTION: of No. 5977336-Typeable Haemophilus
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Shoemaker and Matzare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: STREET: Bldg. 1
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/617,697
: FILING DATE: 01-Apr-1996
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 05-Oct-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry M
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-557
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1599 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-617-697-9
alignment_scores:
Quality: 308.50 Length: 1542
Ratio: 0.439 Gaps: 75

```

Percent Similarity: 45.525 Percent Identity: 20.169

alignment_block:

US-09-303-518D-653 x US-08-617-697-9 ..

Align seg 1/1 to: US-08-617-697-9 from: 1 to: 1599

```

1 ATGAACAACCGCAGCAACGCAACGCAACACACCGCAACGCCCTAA 50
  ||| ||| ||||| ||| |||
327 MetIleThrGlyAspLysValThrLeuLysThrGlyAlaValIleAspLe 343
51 AACCGCGCGATCCGCTTCGCCCGCTTACTTACCATATGCGCTGCTG 100
  ::|||::
343 uSerGlyLys..... 346
101 TCGCATTCGCCCCAAGCCCGCGGAGACACTTATTTGCGCATCAAC 150
  ::|||::|||::|||::|||
347 .....GluGlyGlyLysThrThrLeuGly..... 354
151 TACCAATATATCGCGACTTTGCGCAAAATAAGCAAGTTTGCAGTCG 200
  ::|||::|||::|||::|||
355 .....GlyAspGluArgGlyGlu.....Gly 361
201 GCGCAAGATATGAGTTTACACAACAAAAAGGAGTGTGCGCAAT 250
  ::|||::|||::|||
361 YLysAsnGlyIleGlnLeuAlaLysLys.....ThrT 372
251 CGATGACAGAACCCCGCATGATGATTTCTGTGATGCGGTAACGCG 300
  ::|||::|||::|||
372 hLeuGluLysGlySerThrIleAsnValSerGlyLysGluLysGly 388
301 GTGGCGCATTTGGCGGCGATCAATATATTGTG..... 333
  ||| ::|||
389 ArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAl 405
  ::|||::|||::|||
334 .....AGCGTGACATTAACGCGCG..... 354
  ::|||::|||::|||
405 aglnGlyLysAspIleAlaLysThrGlyGlyPheValGluThrSerGly 422
355 ..TATTAACATGTTGATTTGTCGCGAGGAGCAATGCC..... 393
  ||| ::|||
422 IStyLeuSerIleAspAspAsnAlaIleValLysThrLysGluTrpLeu 438
394 ..GATCAGCACCGCTTTCTTACCAATGTTGAAAAGAAATAATATTA 440
  ||| ::|||
439 LeuAspProGluAsnValThrIleGluAlaProSerAlaSerArgValG 455
441 AGCAGGAGCTAACGCGCATCTTATGCGGCGGATTAATCATATGCCG 490
  ::|||::|||::|||
455 uLeuGlyAlaAspArgAsnSerIleSerAlaGluValIleLysValThr 472
491 TGCACAATTTGTCAAGATGACAGAACCTGTTGAGATGACGAGTAAAT 540
  ||| ||| ||| |||
472 euLysLysAsnAsnThrSerLeuThrThrLeuThrAsnThrThrIleSer 488
541 GATGGGTGGAATAACGCTGATTAATAATATACCTGATCGTGTGCAAT 590
  ::||| ||| ::|||
489 AsnLeuLysSerAlaIleValIleValAsnIleThrAlaArgAlaLys 505
591 C..... 591
505 uThrValAsnSerSerIleSerIleGluArgGlySerHisLeuIleuH 522
592 .....GGACAGAGCAGACAATATGCGGCTGTGATGAGAGCAACCC 633
  ||| ||| ||| ::|||
522 IAspGlnGlyGlnGlyGlnGlyValGlnIleAspLysAsp..... 536
634 AATAACCGCGAAAGTTCAATATATGCAAGCGCATATCTTGCGTCGT 683
  |||::|||
537 .....IleThrSerGlu..... 540
684 CGGTGGCATATCTTTGCACAAAATGATCGAGTGGT..... 720
  |||||

```

```

541 GlyLysAsnLeuThrIleTyr.....SerGlyGlyTrpValAspValH 555
721 ..GGCAGAGTCACCTTAGTAGTAGCGAAATAATTAACATAGCCCATGCT 768
  ::|||::|||::|||
555 IAspAsnIleThrLeuGlySer.....Gly 563
769 TTTTACCAACAGAGAGCTCATTTGGCGAC.....AG 800
  ||||| ::|||
564 PheLeuAsnIleThrThrLysGlnLysPheIleAlaPheGluAspLysSe 580
801 TGGCTCACCAGATGTTATCTATGATGCCCCAAGCAAAAGGTGAATA 850
  ||| ::|||
580 rGlyArgAsnAsnLeuThrIleThrAlaGln..... 590
851 ATGGGATATGCAACAGCAGCAACCCCTATATAGAAAAAGCAATGCTTC 900
  |||::|||::|||
591 ..GlyThrIleThrSerGlyAsn.....SerAsnGlyPhe 601
901 CAGCTAGTTCTGAAGATGTTCTATGATGCAATCTTGTGCGAGATAC 950
  ::|||::|||
602 Arg.....PheAsnAsnValSerLeuAsnSerLe 611
951 CCATTCAGTATCTACGAAACACATCAAAATGCAAAATATCTTTTAACG 1000
  ::|||::|||
611 u.....GlyGlyLysLeuSerPheIleThr 619
1001 ACAATATATATGCGCGAGCAAA..... 1023
  |||::|||
619 sPserArgGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLys 635
1024 .....ATCGATGCCAACA 1037
  |||::|||
636 PheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysAl 652
1038 TAAACACTATCTCTACCTTATGATTAACAGCAACCGTT..... 1080
  ||| ||| ||| |||
652 aProLysValSerTrpPheThrArgAspLysGlyArgThrTrpAsnV 669
1081 .....CAATTGTTAATGTTCTTTATTC 1104
  ::|||::|||::|||
669 alThrThrLeuAsnValThrSerGlySerLysPheAsnLeuSerIleAsp 685
1105 GAGACAGCAGAGAACCTGTTATCATGCTGCAGGTGGGCTAACAGTTA 1154
  ::||| |||
686 SerThr.....GlySerGlySerTh 692
1155 TCGACCCAGACTGATATATGAGAAATATTTCTTTATGCAAAAGNA 1204
  ||| ::|||
692 rGlyProSerIleArgAsnAlaGlu..LeuAsnGlyIleThrPheAsnL 708
1205 AAGGTGAATGATCTTACCAGCAACATCAACAGCGCGCGGCGTTTG 1254
  ||| ::|||
708 ySalA.....ThrPheAsnIleAlaGlnGlySer..... 717
1255 TATTTGAGGATATTTTACGGCTCGCGCT.....AAAA 1289
  ::|||::|||::|||
718 .....ThrAlaAsnPheSerIleLysAlaSerIleMetProPheLysSe 732
1290 CAACGAAACGTGCGCAA.....GCGCGGCGCTTCAATATACGTGATGCA 1333
  ||| ::|||
732 rAsnAlaAsnTrpAlaLeuPheAsnGluAspIleSerValSerGlyGlyG 749
1334 GTACCGTTACTTGAAGATTAACGCG..... 1359
  ::|||::|||::|||
749 ySerValAsnPheLysLeuAsnAlaSerSerSerAsnIleGlnThrPro 765
1360 .....GTGGCAACGACGCGCTGTCCAAATGCGCAAGGACGCGT 1400
  ::|||::|||
766 GlyValIleLeuLysSerGlnAsnPheAsnValSerGlyLysThrIle 782
1401 GCTGTTCAAGCAAGG.....GAAACCC 1426
  |||::|||
782 uAsnLeuLysAlaGlnGlySerThrGluThrAlaPheSerIleGluAsnA 799

```

```

1427 AAGGCTCGCTACGCGTGGCGACGTAAGTCATCTTAGATCAG...CAG 1473
:: :::::::::::::: :::::::::::::: ::::::::::::::
799 sPLeuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGlnValIu 815
1474 GCGGACGATCAA..... 1485
:: ::::::::::
816 GlyThrAspSerArgValAsnGlyValAlaAlaLysLysAsnIleTh 832
1486 ..... GGCACAAAACACGCTTTAGTGAAA 1510
832 rPheLysGlyGlyAsnIleThrPheGlySerGlnLysAlaThrThrGlu 849
1511 TCGGCTTGCTACGCGGAGGAGCGTGTCAATGTAATGCCATATCAG 1560
:: :::::::::::::: :::::::::::::: ::::::::::::::
849 Le.....LysGlyAsnValThrIleAsnLysAsnThrAsn 860
1561 TTCAACCCCGACAAACTATTTGCGCTTTCGCGGCGGACGTTTGATTT 1610
:: :::::::::::::: ::::::::::::::
861 AlaThr.....LeuArgGlyAlaAsnPheAlaG 870
1611 GAACGGGCGATTCGCTTTCGCTTCCACCGCATTCAAATACCATGAGGG 1660
:: ::::::::::::::
870 uAsnLysSerProLeuAsnIle.....AlaGly 880
1661 CGATGATTTGTCAACCCACAAT...CAAGACAAGAATCCACGCTTACATT 1707
:: :::::::::::::: ::::::::::::::
880 snValIleAsnAsnGlyAsnLeuThrThrAlaGlySerIleIleAsnIle 896
1708 ACAGGCATTAAGATATTACTACAAACGCGCAAT..... 1740
:: ::::::::::::::
897 AlaGlyAsnLeuThrValSerLysGlyAlaAsnLeuGlnAlaIleThrAs 913
1741 ..... AACACAACTTGATGATGCAAAA 1762
913 nTyrThrPheAsnValAlaGlySerPheAspAsnGlnLysAlaSerAsn 930
1763 AAGAAATTCCTTACACAGCGTTGGTGGCGAAGAAATGCAACCAAAACG 1812
:: :::::::::::::: ::::::::::::::
930 LeSerIleAla...ArgGlyGlyAlaLysPheLysAspIleAsnAsnThr 945
1813 AACGGGCGCTCATCTGAATTACCAACCGAAGACGGGATCGCATTT 1862
:: :::::::::::::: ::::::::::::::
946 SerSer...LeuAsnIleThrThrAspSerAspThrTyrArgThrI 961
1863 ACTGCTTCCGGCGAACAATTAACGCAATATACGCAACAACAACG 1912
:: ::::::::::::::
961 e.....IleLysGlyAsnIleSerAsnLysSerG 971
1913 GCAAACTGTTTTCAGCGCGACGCGACGCGACGCGCTTACATCATTTA 1962
:: :::::::::::::: ::::::::::::::
971 LysAspLeuAsnIleIleAspLysSer...AspAlaGluIleGlnIle 986
1963 GGAAGCGGCTGTCAAAATGGAAGGATTCGCCAAGAGAGAATCGGTG 2012
:: :::::::::::::: ::::::::::::::
987 GlyGlyAsnIleSerGlnLysGlu.....GlyAsnLeuThrI 999
2013 GGACACGATTTGATCGAC.....CGCACATTTAAACGCGAAA 2050
:: ::::::::::::::
999 eSerSerAspLysValAsnIleThrAsnGlnIleThrIleLysAla... 1014
2051 ACTTCATATTACAGGCGGACACG..... 2076
:: ::::::::::::::
1015 .....GlyValGluGlyGlyArgSerAspSerSerGluAlaGluAsnAla 1029
2077 .....GTGGTTCCCGAATGTTGCCAAGTGAAGGAGCGATTTGCAATT 2120
:: ::::::::::::::
1030 AsnLeuThrIleGlnThrLysGlnLeuLysLeuAlaGlyAspLeuAsnI 1046
2121 AAGCAATCACGCCAAGCATGTTTTCGCTGTCGACCGCATCAAGCCACA 2170
:: ::::::::::::::
1046 eSerGlyPheAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeu 1063

```

```

2171 CAATCTGACACGTTTCGACTGACGCGGCTCTGACAAAGTTGTACGAAA 2220
:: ::::::::::::::
1063 hTrIleGlyAsnAlaSerGly.....GlyAsnAlaAspAlaLys 1075
2221 ACCATTACCGACGATTAAGTATGCTTCATTAGACAGACCGACATCAG 2270
:: ::::::::::::::
1076 LysValThrPheAspLysValLysAspSerLysIleSerThrAsp...GI 1091
2271 AGGCAATGTACGCTTCGCCATCAGCTCATTTAATCTCACAGACTTG 2320
:: ::::::::::::::
1091 yHisAsnValThrLeuAsnSerGlu.....ValI 1101
2321 CCACACTCACGCGCAATCTTAGTGCAGCGGAGAC.....ACGCACTAT 2364
:: ::::::::::::::
1101 yThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerThrGlyLeu 1117
2365 ACGGTTACG...CGCAACGCCACCCAAACAGCGACCTCAGCTCGTGG 2411
:: ::::::::::::::
1118 ThrIleSerAlaLysAspValThrValAsnAsnValThr..... 1131
2412 CAATGCCCAAGCAACATTTATCAAGCCACATTAAACGCAACATCGG 2461
:: ::::::::::::::
1132 ....SerHisLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThr 1147
2462 CTTCGACAAATGCTTCATTATCTAAGCAACAACGCCGTA..... 2502
:: ::::::::::::::
1147 hTrLysGluGlyThrThrIleAsnAlaThrThrIleSerValGluValThr 1163
2503 .....CAAAACGCGCAGCT 2516
1164 AlaGlnAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrVa 1180
2517 GACGCTTTCGACAC.....GCTAAGCAACGTAAGCATTCGCCAC 2560
:: ::::::::::::::
1180 IThrAlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAla 1197
2561 TCAACGCAATGTCTCCCTACGCGATAGCACTATTCATTTTGAAAC 2610
:: ::::::::::::::
1197 hTrSerGlyThrValAsnIleSerThrLys..... 1206
2611 AGCGCTTTAACGGAATAATCAGCGCGCGCAACGATACGGCA..... 2652
:: ::::::::::::::
1207 .....ThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyAs 1220
2653 ..TTAACTTTAAAGACAGCAATGACGCGCTCGCGTGGCGACGATTA 2701
:: ::::::::::::::
1220 nValAsnIleThrAlaSerGlyAsnThrLeu.....LysValS 1233
2702 GCAATTTAAACCTTGACAAACGCCACCATTAACATTCGCGCTATCGA 2751
:: ::::::::::::::
1233 eAsnIleThrGlyGlnAspValThrValThrAlaAsp..... 1245
2752 CACGATGCGCGCGCGCAACCGCGACGTGGCGGACAGTCGCCGCCG 2801
:: ::::::::::::::
1246 .....AlaGlyAlaLeuThrThrThrAlaGly..... 1254
2802 CCGTTCGCGCGCTTCCTATTCGCTTACGCGCGCACTTCGCGCAAT 2851
:: ::::::::::::::
1255 .....SerThrIleSerAlaThrThrGlyAsnAla..... 1264
2852 CCGGTTTCAACACGCTGACGCGTAACGCAATTTGAACGCTGAGGACA 2901
:: ::::::::::::::
1265 .....AsnIleThrThrLysThrGlyAspIleAsnGlyLys..... 1276
2902 TTCGCTTATGTGGAACCTTCGCGCTACCGACGCGCAATTTGAAGCT 2951
1276 ..... 1276
2952 GCGGAAAGTTCGAGGACCTTAC.....ACCT 2980
:: ::::::::::::::
1277 ValGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThrL 1293
2981 TGCGTGTCAACAATACCGGCAACGAACCGTAAGTCTGAGCAATTTGACG 3030

```

```

1538 LeuN1aLysIeuNGlyAlaSerAlaValArgPheValGluProAsnAsnAl 1554
3820 TTGCAGCAGCGGCATCGGCACCTGGCAGCGGCTTGGCCACAGGTCCGCTTT 3865
1554 alleThrrVal.....AsnThrIeuGlnsngIupheThrThrLysProSerS 1569
3870 CGGGCATTACGGCATCGGCAGCGGTTCACAA.....TCGCATCAGCAGCGG 3913
1559 erGlnValThrIleSerGluGlyLysAlaLysCysPheSerSerIylAsnGly 1585
3914 GCAGCGGCTTTTATGACGGCAGGCC 3937
1586 AlaArgValCysThrAsnValAla 1593
seq.name: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:US-08-617-697-10

```

; Patent No. 5977336

APPLICANT: Barenk

TITLE OF INVENTION: OF

CORRESPONDENCE ADDRESS:

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington;

COUNTRY: U.S.A.

DATE: 2202 0200
COMPUTER READABLE FORM:

COMPUTER: IBM PC COMP

```

;
; OPERATING SYSTEM: PC
; SOFTWARE: PatentIn R
;

```

CURRENT APPLICATION DATE:

APPLICATION NUMBER:

CLASSIFICATION: 424
FILING DATE: 01-APR-

PRIOR APPLICATION DATA:

FILING DATE: 05-OCT-1991

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION

REGISTRATION NUMBER:

TELECOMMUNICATION INFORMATION;

TELEFAX: (703) 415-0

SEQUENCE CHARACTERISTIC

TYPE: amino acid

TOPOLOGY: linear

05-06-617-037-10

alignment scores:

Quality:	301.50
Ratio:	0.433

Percent Similarity: 46.220

```
alignment_block:
```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```
178 AATAAGCGAAGTTTCAGTCGGCGGCAAGATATTGAGTTTACACAA 227
||| ||| :||| ||| ||| ||| :|||
199 AsnHsGlyLeuIleThrValGly...LysAspGlySerVal..... 211
228 AAAAGGAGTTGGTCGGCAAAATGATGACGAAGCCCGCATGATGATT 277
:||||:||||| :||| :||| :||| :|||
212 .....AsnLeuIleGlyGlyLysValLysAsnGlu..... 221
278 TTTCTGGCTATCCGCTAACGGC.....GTGGCGCATTTGGCGCGCAT 321
:||||:||||| ||||| :||| :||| :||| :|||
222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237
332 CAA.....TATTGTGGAG 335
||| :||| :||| :||| :|||
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleThrSerIleAl 254
336 CGTGGCAGATTAACGGCGCTATTAACATGTTGATTTGGTCGGAGGAA 385
:||||:||||| :||| :||| :||| :||| :|||
234 AlaProGluAsnGluAlaIleAsnLeuGlyAspIleAlaLysGlyG 271
386 GCAATCCCGATCAGACCGCTTTCTTACCAATTTGAAAGAAATATAT 435
:|||||
271 LysAsn..... 272
436 TATAAGCAGGAGCTAACGCCATCTATGCGCGCATTTATCATATGCC 485
272 ..... 272
486 GCGTTGCACAAATTTGTCACAGATGCAAGAACCTGTTGAGTACAGCTT 535
272 ..... 272
536 ATATGATGGTGGAATACGCTGATTTAATAATACCTGATCGTGT 585
||| :||| :||| :||| :|||
273 .....IleAsn.....Val 275
586 CGAATCGGAGCAGCAGACATATTTGGCGCTGATGAGACGAACCCAA 635
||| :||| :||| :||| :||| :|||
276 ArgAlaIleThrIleArgAsnLysGlyLysLeuSerAlaAspSerVal 292
636 TAACCGGGAAGTTCATATCATATTGCAAGCGCATTTCTTG.....C 679
:||||:||||| :||| :||| :||| :|||
292 rLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyLysAlaG 309
680 TCGTCGGTGCATATCCTTTGACAAATGATGATCAGT...GGTGGCACA 726
:||||:||||| :||| :||| :||| :||| :|||
309 LuIleGlyGlyValIleSerAlaGlnAsnGlnAlaLysGlyLys 325
727 GTCAACTTAGTAGCGAATAATTAACATAGCCCATATGTTTACC 776
:||||:||||| :||| :||| :||| :||| :|||
326 LeuMetIleThrGlyAspLysValThr.....LeuLys 336
777 AACGAGGAGCTCATTTGGCGACAGTGGCTACCAATTTATCATATGATG 826
|||||:||||| :||| :||| :||| :||| :|||
336 sThnGlyAlaValIleAspLeuSerGlyLys..... 346
827 CCCAAAGCAAAAGTGTATTAATGAGGTATTCAAACAGCAACCC 876
:||||:||||| :||| :||| :||| :||| :|||
347 .....GluGlyGlyGluThr 351
877 TATATAGA.....AAAAGCATGGCTTCACGATG 908
|||||:||||| :||| :||| :||| :||| :|||
352 TyrLeuIleGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeuAl 368
909 TCGTAAAGATTGTTCTATGATGAATCTTGTGCGAGATACCATTCAG 958
:||||:||||| :||| :||| :||| :||| :|||
368 AlaLysLys.....ThrThrLeuGluLysGlySerThrIleAsnV 381
959 TATTCTACGACACATCAAAATGGAAA..... 987
||| :||| :||| :||| :||| :|||
381 Al.....SerGlyLysGlyLysGlyLysGlyLysArgAlaIle 391
988 .....TACTTTTTTAACGACATATATATATGCGCAGGAAA 1022
```

```
392 ValTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnLysE 408
:||||:||||| :||| :||| :||| :||| :|||
1023 AATGATGCCCAACATTAACACTATTCTACTTATAGATTAAACAC 1072
:||||:||||| :||| :||| :||| :||| :|||
408 RasPileAlaLysThrGlyGlyPhe.....ValGluThrS 420
1073 GAACCGTTCATTG.....TTAATGTTTCTTATCCGAG 1107
:||||:||||| :||| :||| :||| :||| :|||
420 erGlyHisAspLeuSerIleGlyAspAspValIleValAspAlaLysGlu 436
1108 ACAGCAAGAGAACCT.....GTTATCATGCTGCAGGTGGGT 1145
:||||:||||| :||| :||| :||| :||| :|||
437 TrpLeuLeuAspProAspAspValSerIleGluThrLeuThrSerGlyAr 453
1146 CAACAGTTATCGACCCAGACTGAATATGAGAGAATAATATTCCTTAT 1194
||| :||||:||||| :||| :||| :||| :|||
453 GAsn.....AsnThrGlyGluAsnGlnGlyTyrThrT 464
1195 .....GACAAAGAAAGTGAATGATCTTACACG 1227
:||||:||||| :||| :||| :||| :||| :|||
464 hGlyAspGlyThrLysGluSerProLysGlyAsnSerIleSerLysPro 480
1228 AACATC...AACCAAGCGCGGCGGTGTGATTTGAGGGTAATTTAC 1274
:||||:||||| :||| :||| :||| :||| :|||
481 ThrLeuThrAsnSerThrLeuGluGlnIleLeuArgArgLysSerTyrVa 497
1275 GGTCTGCGCTTAAACACAGAAACGTGGCAGCGCGGCGCTTCATATCA 1324
:||||:||||| :||| :||| :||| :||| :|||
497 LAsnIleThrAlaAsnAsnArgIleTyrValAsnSerIleAsnLeuS 514
1325 GTGATGGCACT...ACCCTTACTGGAAGTAAAGCGCGTG...GCAAC 1368
:||||:||||| :||| :||| :||| :||| :|||
514 eAsnGlySerLeuThrLeuHisThrLysArgAspLysValLysIleAsn 530
1369 GACCGCGCTGCCAAATCGGCAAGGCGACGCTGCTTCAAGCCAAAG 1418
:||||:||||| :||| :||| :||| :||| :|||
531 GlyAspIleThrSerAsnGlnAsnGlyAsnLeuThrIleLysAlaGlySe 547
1419 G.....GAAACCAAGCTCGGTCAAGCTGGGCGACGTAAGTATCT 1462
:||||:||||| :||| :||| :||| :||| :|||
547 rTrpValAspValHisLysAsnIleThrLeuGlyThrGly.....PheL 562
1463 TAGATCAGCAGCGCGCAGAT.....CAAGCAAAAA 1494
:||||:||||| :||| :||| :||| :||| :|||
562 euAsnIleValAlaGlyAspSerValAlaPheGluArgGluGlyAspLys 578
1495 CAAGCCTTAGTGAATCGCTTGTGACGCGCAGGCGGACGTCACACT 1544
:||||:||||| :||| :||| :||| :||| :|||
579 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrVa 595
1545 GAATGCCGATATCAGTTCAACCCGACAAACTATTTGCGCTTGGCG 1594
:||||:||||| :||| :||| :||| :||| :|||
595 LAsnLysAspAspLys.....GlnPheArgPheAsnA 606
1595 GCGCAGCTTGGATTGACGCGGCGCATTCGTTGCTTCACGCAATTCAA 1644
:||||:||||| :||| :||| :||| :||| :|||
606 snValSerLeuAsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln 622
1645 AATACCGATGAAGGGCGATGATTCACACCAATTCAGACAAAGATC 1694
:||||:||||| :||| :||| :||| :||| :|||
623 AsnAsn.....PheThrHisLysPheAspGlyGlu... 632
1695 CACCGTTACATTCACGCAATTAAGATTTACTACACCGGCAATTAACA 1744
:||||:||||| :||| :||| :||| :||| :|||
633 .....IleAsnIleSerGlyIleValThrIleAsnGlnThr..... 644
1745 ACAACTTGGATGCAAAAGAAATGCTTACACAGGTTGTTGGCGAG 1794
:||||:||||| :||| :||| :||| :||| :|||
645 .....ThrLysLysAspValLysTyr.....TrpAsnAlaSer 655
1795 AAAGATGACAAACAAAGCAAGCGGCGCTCAATGTAATTACCAACCGGA 1844
|||||:||||| :||| :||| :||| :||| :|||
```

656 LysaspserlyrTrpAsnValserSerLeuThrLeuAsnThrValGlnly 672
1845 AGAAGCGGATGCGACTTACTGCTTCCGGCGGACAAATTTAAAGGCA 1894
672 spherThrPheValaspsergly..... 682
1895 ATATCAGCAAAACGCAAACTGTTTTCACGGCGAGACGACACCG 1944
683SerAsnGlylnAspleuArgerSerArg..... 692
1945 CAGCCCTACATCATTTAGAGAGCGGGTGGTCAAAAAATGAGAGTATCCC 1994
692 692
1995 ACAAGAGAAATCGTGTGGGACACAGATGTGATCGACCGACATTTAAAG 2044
693ArgSerPheAlaG 697
2045 CGGAAACTTCATATTCAGCGCGGACAGAGCGGTGTTCGCGCATGTT 2094
697 LysAlaHisPheAsnGlylIeGlyGlySerThr..... 707
2095 GCGAAAGTGGAGCGGATGGCATTTAAGCATACGCCACAGCATTTT 2144
708AsnPheAsnIleGlyAlaAsnAlaLysAlaLeuPh 719
2145 CGGTGTGCGACCGCAT..... 2160
719 eLysLeuLysProAsnAlaAlaThrAspProLysGlyLeuProIleT 736
2161CAAAGCCACACATCTGTACAGCTTCGACGCGGCTGTGACA 2205
736 hrPheAsnAlaAsnIleThrAlaThrGlyAsnSerAspSerValMet 752
2206 AGTGTACCGGAAAAACCATTCAGCAGATTAAGTATGATGCTTATTGAG 2255
753 PheAspIleHisAlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetAs 769
2256 CAAAGCCGACATCAGAGCATGTGACGCTTCGCGCATCAGCGCAT..... 2301
769 pSerIleAsnIleThrGlyGlyLeuAspPheSerIleThrSerHisAsn 786
2302TTAATCTGACA 2313
786 rGAsnSerAsnAlaPheGlyIleLysLysAspLeuThrIleAsnAlaThr 802
2314 GGA..... 2316
803 GlySerAsnPheSerLeuLysGlyThrLysAspSerPheTyraAsnGlyTy 819
2317CTTGCCACACTCAACGCA 2335
819 rSerLysHisAlaIleAsnSerSerHisAsnLeuThrIleLeuGlyAla 836
2336 ATCTTAGTGCAGCGGAGACACGACATACGTTACGCGCAACGCCAC 2385
836 snValThrLeuGlyGlyGlu.....AsnSerSerSerSer 847
2386 CAAAACGCGCAACCTCAGCTGTGGGCAATGCCAAGACATTTAAATCA 2435
848 IleThrGlyAsnIleAsnIleThrAsnLysAlaAsnValThrLeuGlyAla 864
2436 AGCCACATTAAACGCGCAACATCGGCTTCGACATGCTTCAATTAATC 2485
864 asPheThrSerAsnSerAsnThrGlyLeuLys...LysArgThrLeuThrL 880
2486 TAAGCAACAGCGCGTACAAAAGCGCATGTGACGCTTCGCAACAGCT 2535
880 euGlyAsnIleSerValGlu...GlyAsnLeuSerLeuThr...GlyAla 894
2536 AAGGCAAACTGAAGCATTCGCGCACTCAACGATGCTCCCTAGCGCA 2585
895 AsnAlaAsnIleVal.....GlyAsnLeuSerIleAlaGly 906
2586 TAAGCAGTATTCATTTGAA.....ACAGCCGCTTACCGGAA 2626
906 uAspSerThrPheLysGlyGlyAlaSerAspAsnLeuAsnIleThrGlyT 923
2627 AATATCAGCGCGCAAGATACGCGATTACACTTAAAGACAGCGCAATGG 2676
923 hrPheThrAsnAsnGlyThrAlaAsnIleAsnIleLys..... 935
2677 AGCGTCCGTCGCGGCAAGTATTAAGCAATTTAAACCTTGACACAGCCAC 2726
936GlyValValLysLeuGlyAspIleAsn..... 944
2727 CATTAACATCAATTCGCGCTATCGACAGATGCGGAGCGGCAACCG 2776
944 944
2777 GCATGCGGACATGCGCGCGCGCGCTTCGCGGCTTCCATTTATTC 2826
945AsnLysGlyLysLeuAsn 950
2827 GTTACGCGCGCAACTTCGCGCAGATCCGTTCAACAGCGTACGCTAA 2876
951 IleThr.....ThrAsnAlaSerGlyThrGlnLysThrIle...IleAs 964
2877 CGGCAAAATG...AACGTCAGGAAACATTCGCTTATGTG..... 2916
964 ngLysAsnIleThrAsnGlyLysGlyAspleuAsnIleLysAsnIleLysA 981
2917GACTCTTCGCTACCGCAGCGGCAATTTGAAGCTG 2952
981 laAspAlaGlyIleGlyIleGlyGlyAsnIleSerGlnLys..... 994
2953 GCGGAAAGTTCCGAAGCACTTACACCTTGCT.....GTCAACA 2993
995GluGlyAsnLeuThrIleSerSerAspLysValAsnI 1007
2994 TACCGGCAACGAACCCGTAGTCTCGAGCAATTCAGGTA.....G 3034
1007 eThr.....AsnGlnIleThrIleLysAlaGly 1017
3035 TGGAA...GGAAAAGACACACAGCGCTTCGGAATTTAATTCACC 3081
1017 alGlyIleGlyLysSerAspSerSerGlyAlaGlyAsnAlaAsnLeuThr 1033
3082 CTCGAAACGACACGTCGATGCGCGCGCATGCTTATCAGCTTATCG 3131
1034 IleGln..... 1035
3132 CAAGAAGCGGAGTTCGCTGCATATATCCGTCGAAGACAGACTTT 3181
1036ThrLysGlyLeuLysLeuA 1042
3182 CCGCAAACTCGGCAAGCGGCGGGAACAGAGGCGCTTGAACGCAAA 3231
1042 laGlyAspleuAsnIleSerGlyPheAsnLysAlaIleThrAlaLys 1058
3232 CAGGCAAACTTCGCGCAACACAGCGGGAAGAACAGACCGCAAG 3281
1059 AsnGlySerAspLeuThrIleGlyAsnAlaSerGlyAlaAsnAlaAspAl 1075
3282 C.....CTTGACCGCTGATTCGGGCG...G 3304
1075 alLysLysValThrPheAspLysValLysAspSerLysIleSerThrAspG 1092
3305 GCGCAATGCGCACGAAAGCGAGAA...AGTGTTCGCAACGCGCGCG 3351
1092 LysHisAsnValThrLeuAsnSerGlyValLysThrSerAsnGlySerSer 1108
3352 CAGCAGAGCGGGAATAATGCC...GGCATTTATGACGCGGAGAGAACG... 3396
1109 AsnAlaGlyAsnAspAsnSerThrGlyLeuThrIleSerAlaLysAspVa 1125

```

3397 .....AAAAAGGGTGCAGCGGATTAAG 3421
1125 lThrValAsnAsnValThrSerHisLysThrIleAsnIleSerAla 1142
3422 ACACCGCTTGGCGAAGACGGCGAAGACCGCGCGCTACCAAC 3471
1142 lAsnIleValThrThrLysGluThrIleAsnAlaThrThr 1158
3472 GCCTTCCCGCGCGCGCGCGGATTTGGCGAAGCGGACCC 3521
1159 Gly.SerValGluValThrAlaGlnAsnGlyThrIleLysGlyAsnIleT 1175
3522 CCAACCGCAACCCCAACCGCAGCGCGATCATCAGCGCTATGCGCAATA 3571
1175 hrSerGlnAsnValThrValThrAlaThrGlnAsnLeuValThrThrGlu 1191
3572 GCGGTTGAGTATTTCCGCCCAACGCTCAACAGCGTTTCCGCTACAG 3621
1192 AsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer..... 1205
3622 GACGAATTTGACCGCGCTTTTCCGCAAGACCGCGCAACCGCTTGA. 3670
1206 .....ThrLysThrGlyAspIleLysGlyIleT 1215
3671 ..CAAGCGCATTCGGGACACCAACACTACCGTTCG..... 3706
1215 lGluSerThrSerGlyAsnValAsnIleThrAlaSerGlyAsnThrLeu 1231
3707 AAGATTCCCGCGCTACCGCAACACCGCGCGCAATCGGATG 3756
1232 LysValSerAsnIleThrGlyGlnAspValThrValThrAlaAspAlaG1 1248
3757 CAGAAACCTCGGCGACGGCGCGGCGGATCTGTTTTCGCAACACG 3806
1248 yAlaLeuThrThrAlaGlySerThrIleSerAla.....ThrThrG 1263
3807 GA..... 3808
1263 lYsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValGlu 1279
3809 CCGGAACACCTTCGACGACGCGATCGCAACTCGGACGCGTTGGCCAC 3858
1280 SerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAlaVa 1296
3859 GGTGCGCTTTTCGGGCAAT.....ACGCGATCGGCGAGTTTGA 3896
1296 lGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGlyLysL 1313
3897 CATCGGATCAGCGCGGCGGCTTTTACTAGCGGACGCTTTCAGACG 3946
1313 eutThrSerThrValGlySerThrIleAsnGlyThrAsnSerValThrThr 1329
3947 GCATCAGAGCAAAA.....TCCGCGCGCGCGTGC 3978
1330 SerSerSerSerGlyAspIleGluGlyThrIleSerGlyAsnThrValAs 1346
3979 CATTCAGCGCATTAAGCGAAGAT.....ACGCGCGAGTTT 4013
1346 nValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAlaLysValG 1363
4014 CGGCGGATTCGGCATCGAACCACATCGGCGCAACGCGTATTTGCTGC 4063
1363 lValAlaLysAsnGlyAlaAlaThrLeuThrAlaGlu.....Ser 1375
4064 AAAAAGCGGATTCAGTACGAAAGATCATATCGCAACCGCGGCGCTT 4113
1376 GlyLysLeuThrThrGlnThrGlySerSerIleThrSerSerAsnGly.. 1391
4114 GCATTCACCGCTACCGCGCGG 4135
1392 .GlnThrThrLeuThrAlaLys 1398
seq_name: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.us-08-728-470-10

```

```

seq_documentation_block:
/ Sequence 10, Application US/08728470
/ Patent No. 5928651
/ GENERAL INFORMATION:
/ APPLICANT: Barenkamp, Stephen J
/ TITLE OF INVENTION: High Molecular Weight Surface Proteins
/ TITLE OF INVENTION: of No. 5928651-Typeable Haemophilus
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Shoemaker and Matlare, Ltd.
/ STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
/ CITY: Arlington
/ STATE: Virginia
/ COUNTRY: U.S.A.
/ ZIP: 22202-0286
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentln Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/728,470
/ FILING DATE:
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/302,832
/ FILING DATE: 16-MAR-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US PCT/US93/02166
/ FILING DATE: 16-MAR-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: GB 9205704.1
/ FILING DATE: 16-MAR-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Berkstresser, Jerry W
/ REGISTRATION NUMBER: 22,651
/ REFERENCE/DOCKET NUMBER: 1038-633
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 415-0810
/ TELEFAX: (703) 415-0813
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1529 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ US-08-728-470-10

alignment_scores:
Quality: 292.50 Length: 1494
Ratio: 0.415 Gaps: 77
Percent Similarity: 47.189 Percent Identity: 20.415

alignment_block:
US-09-303-518D-653 x US-08-728-470-10 ..
Align seg 1/1 to: US-08-728-470-10 from: 1 to: 1529

178 AATAAAGCAAGTTGACGCGGCGCAAGATATTTAGTTTACACAA 227
117 lll .....lllll .....lllll
127 AsnHisGlyLeuIleThrValGly..LysAspGlySerVal..... 139
228 AAAAGGAGTTGTGCGCAATCGATGACGAAGACCGCGATGATTGATT 277
140 .....AsnLeuIleGlyGlyLysValLysAsnGlu..... 149
278 TTTCGTGTATCGGTAAGGC.....GTGCGCGCATTTGCGGCGGAT 321
150 ..GlyValIleSerValAsnGlyGlySerIleSerLeuValAlaGlyGln 165

```

```
322 CAA.....TATATGTGAG 335
    :|||
166 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIleIle 182
    :|||
336 CGTGGCAGTAAACGCGGCTATTAACAATGTTGATTTGGTCGAGGAGAA 385
    :|||
182 aaIaProGluAsnGluAlaIleAsnLeuGlyAspIlePheIleAlaGlyG 199
386 GCAATCCCGATCAGCACCGCTTTCTTACCAATTTGTGAAGAATAAT 435
    :|||
199 Lysn..... 200
436 TATAAACGAGGAGTAACGGCCATCCTTATGGGGGATTAATATATGCC 485
200 ..... 200
486 GCGTTTCACAAATTTGTACAGATGCAGAACCTGTTGAGATGACAGTT 535
200 ..... 200
536 ATATGATGGGTGGAATACGCTGATTTAATAATAATACCTGATCGTT 585
    :|||
201 .....Ileasn.....Val 203
586 CGAATCGAGAGCGACACATATATGCGCGCTGATGAGACGACGACCA 635
    :|||
204 ArgAlaIleThrIleArgAsnIleGlyLeuSerAlaAspSerValSe 220
636 TAACCGGGAAGTTCAATATCATATTTGCACGCGATATCTTGG.....C 679
    :|||
220 rIyAspIleSerIleAsnIleValIleSerAlaIleGlyIleValIle 237
680 TCGTCGCTGGCAATCCTTTCACAAATATGATCAGT...GGTGGACAC 726
    :|||
237 IuIleGlyIleValIleSerAlaIleAsnGluAlaIleGlyGlyLys 253
727 GTCAACTAGTAGCGAATAATTAACATAGCCCATATGTTTATACC 776
    :|||
254 LeuMetIleThrIleGlyAspIleValThr.....LeuLys 264
777 AACGAGAGCTCATTTGGCGACAGCTGCTCACCATGTTATATGATG 826
    :|||
264 sThrGlyValValIleAspLeuSerIleLys..... 274
827 CCCAAAGCAAAAGTGTATTAATGAGGATGCAACAGCAACCC 876
    :|||
275 .....GluGlyGlyGluThr 279
877 TATATAGGA.....AAAGCAATGCTTCACAGTAGT 908
    :|||
280 TyrLeuGlyIleAspGluArgIleGlyLysAsnGlyIleGluIleVal 296
909 TCGTAAGATGTGTTCTATGATGAATCTTTCGTCGAGATACCATTCAG 958
    :|||
296 aLysLys.....ThrThrLeuGluLysGlySerThrIleAsnV 309
959 TATCTACGACACATCAATAATGGAAA..... 987
    :|||
309 al.....SerGlyLysGluLysGlyArgAlaIle 319
988 .....TACTTTTAAACGACATATATTAATGGCGACGAGAA 1022
    :|||
320 ValThrGlyAspIleAlaIleValIleAspIleAsnIleAsnIleGly 336
1023 AATGATGCCAAACATAACACTATCTCTACCTTATAGATTAATAACAC 1072
    :|||
336 rAspIleAlaLysThrGlyGlyPhe.....ValGluThrS 348
1073 GAACCGTTCAATG.....TTTAATGTTTCTTTATCCGAG 1107
    :|||
348 ergIyHisAspLeuSerIleGlyAspSerValIleValAspAlaLysGlu 364
1108 ACAGCAAGAGAACT.....GTTTATCATGCTGCAGAGTGGGGT 1145
```

```
365 tryPleuLeuAspProAspAspValSerIleGluThrIleThrSerGlyAr 381
    :|||
1146 CAACAGTTATGACCCAGACTGATTAATGAGAAATATTTCTTTAT 1194
    :|||
381 gAsn.....AsnThrGlyIleAsnGluIleYThrT 392
1195 .....GACAAAGGAAAGGTAATGATGATTAACAGC 1227
392 hGlyAspGlyThrIleLysGluSerProLysGlyAsnSerIleSerLysPro 408
1228 AACATC...AACCAAGCGCGCGGCTTTGTAATTTGAGGTAATTTAC 1274
    :|||
409 ThrLeuThrAsnSerThrLeuGluGluIleLeuAlaArgIleGlySerTy 425
1275 GGTTCGCTTAAACAAACAGACGTCGCAAGCGCGGCTTCATATCA 1324
    :|||
425 IAsnIleThrAlaAsnAsnArgIleTyValAsnSerIleAsnLeuS 442
1325 GTGATGGCAGT...ACGTTACTTGAAGTAACGCGGTG...GCAAC 1368
    :|||
442 eAsnGlySerIleuThrIleuHisThrLysArgAspGlyValLysIleAs 458
459 GlyAspIleThrSerAsnGluAsnIleLysLeuThrIleLysAlaGlySe 475
1419 G.....GAAACCAAGCGCTCGTCACGCGTGTTCAGCCCAAG 1418
    :|||
475 rIrrValAspValHisLysAsnIleThrLeuGlyThrGly.....PheL 490
1463 TAGATCAGCAGCGCGAGCAT.....CAAGCGCAAAA 1494
    :|||
490 euAsnIleValAlaGlyAspSerValAlaPheGluArgIleGlyLysP 506
1495 CAAGCCTTATGTAATGCGCTTGTGTCAGCGCGAGGAGCGGTCACT 1544
    :|||
507 AlaArgAsnAlaThrAspAlaGluIleThrAlaGluGlyThrIleThrVa 523
1545 GAATGCCGATTAACGATTCACCCGACAAACTATTTGCGCTTCGCG 1594
    :|||
523 IAsnLysAspAspLys.....GlnPheArgPheAsnA 534
1595 GCGGAGCTTGGATTTGAACGGGCAATGCTTTCCTCCACCGCACTTAA 1644
    :|||
534 snValSerIleAsnGlyThrGlyLysGlyLeuLysPheIleAlaAsnGln 550
1645 AATACCGATGAAGGGCGATGATGTGCAACACATCAAGACAAAGAATC 1694
    :|||
551 AsnAsn.....PheThrHisLysPheAspGlyGlu.. 560
1695 CACCGTTACATTAACGCAATTAAGATTAATTAACACCGGCAATTAACA 1744
    :|||
561 ...LeuAsnIleSerGlyIleValThrIleAsnGlnThr..... 572
1745 ACAACTGTGATAGCAAAAAGAAATTTGCTACACGTTGTTGGCGAG 1794
    :|||
573 .....ThrLysLysAspValLysTy.....TyrAsnIleAsn 583
1795 AAGATGCAACCAAGCAAGGAGCGGCTCAATCTG..... 1830
    :|||
584 LysAspSerTyThrAsnValSerSerLeuThrIleAsnThrValGlnGly 600
1831 .....AATTAACAACCGGAAG 1846
600 sPheThrPheIleLysPheValAspSerGlySerAsnSerGlnAspLeuA 617
1847 AAGCGATGCGACTTACTG.....CTTCCCGC 1875
    :|||
617 rSerSerIrrArgSerPheAlaGlyValHisPheAsnGlyIleGlyGly 633
1876 GGAACAATTAATTAACGCAATATACGCAACAAACGCAAACTGTTT 1925
    :|||
```


634 LysThrAsnPheAsn.....IleGlyAlaAsnAlaLysAlaLeuPh 647
1926 CAGCGGCGACCGACGA.....CCGACG 1948
647 elysLeuLysProAsnAlaAlaThrAspProLysLysGluLeuProIleT 664
1949 CCTACCAATTCATTTAGGAAGCGGGTGTCAAAATAGAGATATCCACAA 1998
664 hrPheAsn.....AlaAsnIleThrAlaThrGlyAsnSerAsp 676
1999 GGAAATTCGTGTGGACACAGCATTCGACCGCCACATTTAAAGC.. 2046
677 SerSerValMetPheAspIleHis..AlaAsnLeuThrSerAlaAla 692
2047GAAACTTCATATTCAGCGCGGACAAAGCGGTGTTT 2083
692 aGlyIleAsnMetAspSerIleAsnIleThrGlyLeu..... 705
2084 CCCGCAATGTTGCCAAAGTGGAGCGCATTTAGCAATCAGGCC 2133
706AspPheSerIleThrSerHisAsn 713
2134 CAACGAGTTTCGGTGTCCGACCCATCAAAAGCCACACATCTGTACAG 2183
714 Arg.....AsnSerAsnAlaPheGluIle 722
2184 TTCGACTGGACGGGTCTGACCAAGTTGTACGGAATAAACCAT..A 2227
722 slyAspLeuThrIleAsnAlaThrGlySerAsnPheSerLeuLysGlnT 739
2228 CCGACGATAAAGTATGCTTCATTTGACGAACGACATCAGAGGCAT 2277
739 hrLysAspSerPheTyrAsnGlnLysSerLys..... 749
2278 GTCAGCCTGCGGATCAGCCTCATTTAAATCTACAGAGCTTGCCACT 2327
750HisAlaIleAsnSerSerHisAsnLeuThrIle 761
2328 CAACGCAATCTTAGTCAGCGGAGGACAGCACTATACGGTTACGGCA 2377
761 uGlyGlyAsnValThrLeuGlyGlyGlu.....AsnSerS 773
2378 ACGCCACCAAAACGCGACCTCAGCCTCGTGGCAATGCCAACACAA 2427
773 erSerSerIleThrGlyAsnIleAsnIleThrAsnLysAlaAsnValThr 789
2428 TTTAATCAAGCCATTTAAACGGACACATCGGCTTCGACATGCTTC 2477
790 LeuGlnAlaAspThrSerAsnSerAsnThrGlyLeuLys..LysArgTh 805
2478 ATTAAATCTAAGCAACAACGCCGTACAAACGGCAGTGCAGCTTCCG 2527
805 rLeuThrLeuGlyAsnIleSerValGlu..GlyAsnLeuSerLeuThr. 820
2528 ACAACGCTAAGCAACGATTAAGCATTCGCCACTCAACGCAATGCTTC 2577
821 ..GlyAlaAsnAlaAsnIleVal.....GlyAsnLeuSer 831
2578 CTAGCCGATTAAGCAGATTTCCATTTGAA.....AACAGCGGCTT 2618
832 IleAlaGluAspSerThrPheLysGlyAlaIleSerAspAsnLeuAsn 848
2619 TACCGGAAATATCAGCGCGGACGAAGATAGCGCATTAACCTTAAAGACA 2668
848 eThrGlyThrPheThrAsn..... 854
2669 GCGAATGAGAGCTCCGTCGCGACGGAATTAGCAATTTAAACCTTGAC 2718
855AsnGlyThr.....AlaAsnIleAsnIleLys 863
2719 AAGCCCACTTACACTCAATTCGCTATGACAGACATGCGGAGGCGC 2768
864 GlnGlyValValLysLeuGlnGlyAspIleAsnAsnLysGlyGly..... 878
2769 GCAACCGGAGATGCGGAGATGCGCGCGCGCTTGGCCGCTTCC 2818
878 878
2819 TATTATCGTTACGCCCACTTGCAGAAATCCGTTTCAACAGCGTG 2868
879 ..LeuAsnIleThr.....ThrAsnAlaSerGlyThrGlnLysThrIle 892
2869 ACCGTTAAGCGCAAAATTG..AACGTCAGGAGACATTCGCTTATGTC 2915
893 ..IleAsnGlyAsnIleThrAsnGlnLysGlyAspLeuAsnIleLysAs 908
2916 G.....GAACCTTCGCTACCGCAGCGCAAT 2944
908 nIleLysAlaAspAlaGluIleGlnIleGlyLysAsnIleSerGlnLys. 924
2945 TGAAGCTGGCGGAAGTTCCGAAGGCACTTACACTTGAGCT..... 2985
925GluGlyAsnLeuThrIleSerSerAspLys 934
2986 GTCACAAATACCGGACAGAACCCGTAGTCTGACAGCAATTAGCGTA.. 3033
935 ValAsnIleThr.....AsnGlnIleThrIleLys 944
3034GTGGAA..GGAAGACACACACCGCTGTCCGAATCTTA 3073
944 salagIyValGluGlyGlyArgSerAspSerSerGlnAlaGluAsnAla 961
3074 ATTTCACCTTGCMAAAGCAACACGTCGATGCGGCGCATGGCTTATCAG 3123
961 snLeuThrIleGln..... 965
3124 CTTATCCGAAAGACGGCGAGTTCGCGCTGCATTAATCCGTCAGAAACA 3173
966ThrLysGluLe 969
3174 AGAGCTTTCGCAACAACTGCGCAGCGGAGAAACAGAGCGGCTTGA 3223
969 uLysLeuAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGluIleT 986
3224 CCGCAAAACAGGACACAACTTGCCTCAACACAGCGGAAAGAAACAAC 3273
986 hrAlaLysAsnGlySerAspLeuThrIleGlyAsnAlaSerGlyLysAsn 1002
3274 GCGCAAAAGC.....CTTGACGCGCTGATTCG 3299
1003 AlaAspAlaLysLysValThrPheAspLysValLysAspSerLysIleSe 1019
3300 GGCG...GGCGCAATGCCACGGAAGGACAGA..AGTGTGCGGCAAC 3343
1019 rThrAspGlyHisAsnValThrLeuAsnSerGlnValLysThrSerAsnG 1036
3344 CCGCCCGCAGCAGCGGCGGAAATGCC...GGCATTAAGCAGCGGAG 3390
1036 LysSerSerAsnIleGlyAsnAspAsnSerThrGlyLeuThrIleSerAla 1052
3391 GAAGAG.....AAAAACGGGTGACGC 3413
1053 LysAspValThrValAsnAsnAsnValThrSerHisLysThrIleAsn 1069
3414 GGATAAGACACCGCTTGGCAAAACAGCGGACGGAACCGCGCGG 3463
1069 eSerAlaAlaIleGlyAsnValThrThrLysGlnGlyThrThrIleAsn 1086
3464 CTACCAACGCTTCCCGCGCGCGCGCGGATTTGCGGAA 3513
1086 lAThrThrLys..SerValGluValThrAlaGlnAsnGlyThrIleLysG 1102
3514 CCGCAGCCCAACGCAACCCCAACGCGACGAGCTGATCAGCGGTTA 3563
1102 yAsnIleThrSerGlnAsnValThrValThrAlaThrGlnLysLeuValT 1119

```

3564 TGCCAAATGACGGTTGATGATTTTCGGCCAGCTCAACAGCGTTTCG 3613
      :::::|||||
1119 hrThgluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 1135
      :::::|||||
3614 CCGTACAGACGAATGTGACCGCGTGTTCGCGAAGCCGCCGCAACGCC 3663
      :::::|||||
1136 .....ThrystrhglyAspIlely 1142
3664 GTTTGA...CAAGGGGATCGGGGACCAACACACTACCGTTCCG... 3706
      ||| :::::|||||
1142 sglyeIlyleGuserThrSerGlyAsnValAsnIleThrIleSerGlyA 1159
      :::::|||||
3707 .....AGATTTCCGCGCTACCGCCCAACAAACGCGCTCGCCAAA 3748
      ||| :::::|||||
1159 snThleuLyValSerAsnIleThrglyGlnAspValThValThAla 1175
      :::::|||||
3749 TCGGTATGCAAAAACCTCGGACAGCGCGCTCGGATCTCTTTTCG 3798
      :::::|||||
1176 AspAlaGlyAlaLeuThrThrAlaGlySerThrIleSerAla.... 1190
3799 CACAACCGGA..... 3808
1191 .ThrThrGlyAsnAlaAsnIleThrThrLyThGlyAspIleAsnGlyL 1207
3809 .....CCGAAACACCTTCGACGACGCGCATCGCACTCGCACGCC 3850
      :::::|||||
1207 ySvalGluSerSerSerGlySerValThrLeuValAlaThrGlyAlaThr 1223
3851 TTGCCACAGGTGCGCTTTTCGGCAAT.....ACGCATCGGC 3888
      ||| :::::|||||
1224 LeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSe 1240
3889 AGGTTGCATCGGCATCGACGCGCGCGGCTTTAGTAGCGGACGCT 3938
      :::::|||||
1240 rGlyLySerThrSerThrValGlySerThrIleAsnGlyThrAsnSerV 1257
3939 TTCAGACGCGCATCGACGCAAAA.....TCGCGCGCC 3970
      ||| :::::|||||
1257 alThrThrSerSerGlnSerGlyAspIleGlyThrIleSerIleAsn 1273
3971 GCGTGTGCTGATTCAGCGCATTCAGCGCAAGT.....ACGCG 4005
      :::::|||||
1274 ThrValAsnValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAl 1290
4006 GCAGGTTTCGGGATTCGGGATTCGACGACGCAATCGCGCGCAACGCTA 4055
      :::::|||||
1290 alYsValGluAlaLyAsnGlyAlaAlaThrLeuThrAlaGlu..... 1304
4056 TTTCTCCAAAAAGCGGATTCACCGATACGAAGAAGCTCAATATCGCCACC 4105
      ||| ||| :::::|||||
1305 .....SerGlyLySerThrThrGlnThrGlySerSerIleThrSer 1319
4106 CGGGCTTGATTCACCGCTACCGCGCGG 4135
1320 AsnGly...GlnThrThrLeuThrAlaLyS 1328
seq_name: /cgn2/ptodata/1/iaa/5B_COMB.pep:us-08-719-641-10
seq_documentation_block:
; Sequence 10. Application US/08719641
; Patent No. 6218141
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; TITLE OF INVENTION: of No. 6218141-Typeable Haemophilus
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matzare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.

```

```

; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,641
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832
; FILING DATE: 16-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-625
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1529 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-719-641-10

alignment_scores:
  Quality: 292.50      Length: 1494
  Ratio: 0.415         Gaps: 77
  Percent Similarity: 47.189   Percent Identity: 20.415

alignment_block:
US-09-303-518d-653 x US-08-719-641-10 ..
Align seg 1/1 to: US-08-719-641-10 from: 1 to: 1529

178 AATAAGGCAAGTTTCAGTCGGGCGCAAGATATTGAGTTTACAACA 227
    ||| ||| :::::||||| ||||| :::::|||||
127 AsnHieGlyLeuIleThrValGly...LysAspGlySerVal..... 139
228 AAAAGGGAGTTGGTCGGCAATCGATGACGAAGCCCGGATGATTGAT 277
    :::::||||| :::::|||||
140 .....AsnLeuIleGlyGlyValLysAsnGlu..... 149
278 TTTCTGTGATTCGCGTAAACGCG.....GTGCGGCTATGGGCGGCGAT 321
    :::::||||| ||||| :::::|||||
150 .GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 165
322 CAA.....TATATTGTGAG 335
    :::::|||||
166 LysIleThrIleSerAspIleIleAsnProThrIleThrTySerIleAl 182
336 CGTGCAATACGCGCGCTATACAAATGTTGTTGGTCGGAGAGGA 385
    :::::||||| :::::|||||
182 aaIaProGluAsnGluAlaIleAsnLeuGlyAspIlePheAlaLySgLyG 199
386 GCAATCCCGATGACGACCGCTTTCTTACCAATTTGAAAGAATAATAT 435
    :::::|||||
199 LysAsn..... 200
436 TATAAGCAGGACGTAACGCGCATCTTATGCGGCGGATTAATATATGCC 485
200 ..... 200

```


692 aglylleasmetaspserlleasnillethrglyleu..... 705
2084 CCGCAATGTTCGCAAAAGTGGAGGCGATTTAGCAATACAGCC 2133
706AspSerIleThrSerHisAsn 713
2134 CAAGCAGTTTCGGTGTGCGACCCGCAACAAAGCACAACATCTAGAC 2183
714 Arg.....AsnSerAsnIlePheGluIle 722
2184 TTCGACGAGACGGGTCTGACAGTTGACGAAAAACAT.....A 2227
722 slyspseuthrilleasnallethrglyserAsnPheserleuysglnt 739
2228 CCGAGATTAAGTAGATTGCTTCATTGAGCAAGCCGACATCAGAGCAAT 2277
739 hrllyspserPheIlyrAsnIleuIlySerIys..... 749
2278 GTCAACCTTGGCGATCAGCTCATTTAAATCTACAGGACTTGGCACT 2327
750HisAlaIleAsnSerSerHisAsnLeuThrIle 761
2328 CAAGCGCATCTTAGTCAGGCGGAGACAGCCACTATACGCTTACGGCA 2377
761 uGlyGlyAsnValThrLeuGlyGly.....AsnSer 773
2378 ACAGCACCACAAAGCGCAACCTCAGCTGTGGCAATGCCAAGCAACA 2427
773 erSerIleThrGlyAsnIleasnillethrasnlyslalaasnValThr 789
2428 TTTATACAGCCACATTAACGGCAACACATCGCTTCGCAATGCTTC 2477
790 leuGlnIleAspThrSerAsnSerAsnThrGlyLeuIys...LysArgTh 805
2478 ATTTAATCTTAAGCAACAACGCGTACAAAGCGAGTCTGACGCTTCG 2527
805 rleuThrleuGlyAsnIleSerValGlu...GlyAsnLeuSerleuThr 820
2528 ACAACGCTTAAGCAACGTAAGCCATTCGCACTCAACGCAATGCTTC 2577
821 ..GlyAlaAsnAlaAsnIleVal.....GlyAsnLeuSer 831
2578 CTACCCGATTAAGCGATTTCCATTGAA.....AACAGCGCTT 2618
832 IleAlaGlnAspSerThrPheIysGlyAlaSerAspAsnLeuAsnIle 848
2619 TACGGAATAATCAGCGCGCAAGGATACGCTTAAAGACA 2668
848 eThrGlyThrPheThrAsn..... 854
2669 GCGAATGACGCTGCGCTGCGGACAGGAATTAAGCAATTAACCTTGAC 2718
855AsnGlyThr.....AlaAsnIleasnIleIys 863
2719 AACGCCACCATTAACATTCGCTATCGACAGCAGCGGAGCGC 2768
864 GlnIleValIleValIysLeuGlnIleAspIleAsnAsnIysGly 878
2769 GCAAAACGCGAGTCGCGCAGATGGCGCGCGCGCTTCGCCGTTCC 2818
878 878
2819 TATTATCGTTAGCGCGCAACTTGGCAATCCGTTTCAACAGCGT 2868
879 ..LeuAsnIleThr.....ThrAsnAlaSerGlyThrGlnIysThrIle 892
2869 ACGGTAACGCGCAATG...AACGGTCAGGAGCACTTCGCTTATGTC 2915
893 ...IleasnIlyAsnIleThrAsnGlnIysGlyAspLeuAsnIleIysAs 908
2916 G.....GAACCTTCGCGCTACCGGAGCGGCAAT 2944
908 nIleIysAlaAspAlaGluIleGlnIleGlyIysAsnIleSerGlnIys 924
2945 TGAAGCTGGGGAAAAGTTCCGAAGGCATTACACCTTGCT..... 2985
925GluGlyAsnLeuThrIleSerSerAspIys 934
2986 GTCAACATACCGGCAACGACACCGTAAGTCTGACGCAATTAAGCGTA.. 3033
935 ValAsnIleThr.....AsnGlnIleThrIle 944
3034GTGAA...GGAAGACACACACCGCTGTCCGAAATCTTA 3073
944 salagIyValIleGluGlyIysArgSerAspSerSerGlnIleGlnAla 961
3074 ATTTCACCTGCAAAACGAAACACGTCGATGCGCGGCGATGCTTATCAG 3123
961 snleuThrIleGln..... 965
3124 CTATTCGCAAAAGCGGAGTTCGCTGATTAATCCGTCGAAGACA 3173
966ThrIysGluLe 969
3174 AGAGCTTTCGCAAACTCGGCAAGCGGAGAAACAGAGCGCTTGA 3223
969 uIysLeuAlaGlyAspLeuAsnIleSerIlyPheAsnIysAlaGluIle 986
3224 CCGCAAAACAGGACACAACTTGCCTGCAACAGAGCGGAAAGACAC 3273
986 hralaIysAsnIleSerAspLeuThrIleGlyAsnAlaSerIlyGlyAsn 1002
3274 GCGCAAGC.....CTTGACGCGCTGATTC 3299
1003 AlaAspAlaIysIleValThrPheAspIysValIysAspSerIlyIle 1019
3300 GGCG...GGCGCAATGCCACGCAAAAGGCAAGA...AGTGTTCGCGAC 3343
1019 rThrAspGlyHisAsnValThrLeuAsnSerIleValIysThrSerAsn 1036
3344 CCGCCGCGAGCGAGCGGGGAAATGCG...GGCATTACGACGCGAG 3390
1036 IySerSerAsnIleGlyAsnAspAsnSerThrGlyLeuThrIleSerAla 1052
3391 GAAGC.....AAAAACGGGTGCAGC 3413
1053 LysAspValThrValAsnAsnAsnValThrSerHisIysThrIleAsnIle 1069
3414 GGATTAAGACACCGCTGCGCAACAGCGGAGGAACCGCGCGG 3463
1069 eserAlaAlaIleGlyAsnValThrThrIysGlyIleGlyThrThrIleAsn 1086
3464 CTACACGCGCTTCCCGCGCGCGCGCGCGCGCGGATTTGCCGCA 3513
1086 IeThrThrIleSerValIleValThrAlaGlnAsnGlyThrIleIysGly 1102
3514 CCGCAGCCCAACCGCAACCCACGAGCGGCGCTGATCAGCGTTA 3563
1102 yAsnIleThrSerGlnAsnValThrValThrAlaThrIleAsnLeuVal 1119
3564 TGCATATGCGGTTGAGGAATTTTCGCGCAGCGTCAACAGCGTTTCG 3613
1119 hThrGluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 1135
3614 CCGTACAGGACGAATTAAGACCGCTGTTGCCGAAGACCGCGCAACGC 3663
1136ThrIysThrGlyAspIleIle 1142
3664 GTTTGA...CAAGCGCATCGGAGACCAACAACTACCTTGC... 3706
1142 sGlyIleIleGlnIleThrSerThrGlyAsnValAsnIleThrAlaSerGly 1159
3707AAGATTTCGCGCGCTACCGGCAACAAACGACCTGGCGCAA 3748
1159 snThrLeuIysValSerAsnIleThrGlyIleAspValThrValThrAla 1175

```

3749 TCGGTATCGAGAAAAACCTCGCAGCGGCGCGTCGCGCATCTGTTTCG 3798
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1176 AspaIaegIaIaLeuThrThrThrAlaIaGlySerThrIleSeraIa.... 1190
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3799 CACAACCGGA..... 3808
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1191 .ThrThrGlySnaIaAsnIleThrThrThrGlyAspIleAsnGlyL 1207
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3809 ..... CCGGAACACCTTCGACGACGCGCATCGGCACACGCGCAGCGC 3850
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1207 ysvaIgluSerSerSerIysSerValThrIleValAlaThrGlyAlaThr 1223
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3851 TTGCCACGCGTCCGTTTCGGCAAT.....ACGCGATCGGC 3888
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1224 LeuAlaValIglYasnlIleSerGlyAsnThrValThrIleThrAlaAspSe 1240
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3889 AGTTTCGACATCGGCATCGCGCGCGCGGCGGTTTAACTAGCGCGCAGCT 3938
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1240 rGlyIysLeuThrSerThrValIglYSerThrIleAsnGlyThrAsnSerV 1257
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3939 TTCAGCGCGCATCGAGCGCAAA.....TCGCGCGCC 3970
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1257 alThrThrSerSerIleSerIleAspIleGlyThrIleSerGlyAsn 1273
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3971 GCGTCTCATTCAGCGCATTCAGCGCAAGAT.....ACGCG 4005
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1274 ThrValaIaValThrAlaSerThrGlyAspLeuThrIleGlyAsnSerAl 1290
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4006 GCAGGTTTCGCGGATTCGCGCATCGAACCGCATCGCGCGCAGCGCTA 4055
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1290 alyValaIaValaIaIysAsnGlyAlaAlaThrLeuThrAlaGlu..... 1304
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4056 TTTCGTCGCAAAAGCGGATTCAGCGCATCGCAATATCGCCACCC 4105
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1305 ....SerGlyIysLeuThrThrGlnThrGlySerSerIleThrSer 1319
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1320 AsnGly...GlnThrThrLeuThrAlaIys 1328
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/ptodata1/1aa/bb_COMB.pep:US-08-460-269C-6

seq_documentation_block:
: Sequence 6, Application US/08460269C
: Patent No. 6197548
: GENERAL INFORMATION:
: APPLICANT: CLARE, JEFFREY J.
: ROMANOS, MICHAEL A.
: TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN
: YEAST
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Millen, White, Zelano & Branigan, P.C.
: STREET: 2200 Clarendon Blvd., Suite 1400
: CITY: ARLINGTON
: STATE: VA
: COUNTRY: USA
: ZIP: 22201
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460,269C
: FILING DATE: 02-Jun-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Lebovitz, Richard M.
: REGISTRATION NUMBER: 37,067
: REFERENCE/DOCKET NUMBER: Popov-2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 243-6333
: TELEFAX: (703) 243-6410

```

```

: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 922 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-08-460-269C-6

alignment_scores:
      Quality: 275.50      Length: 1162
      Ratio: 0.566      Gaps: 54
      Percent Similarity: 41.910      Percent Identity: 21.687

alignment_block:
US-09-303-518d-653 x US-08-460-269C-6 ..
Align seg 1/1 to: US-08-460-269C-6 from: 1 to: 922

1366 AGACGCGCCTTCGCAAAATCGCGCAAGG.....ACGCT 1400
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 AsmSerIeuSerIeuArgIleValIysAlaAlaProIeuArgThrTh 18
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1401 GCGTTCAGCCCAAGG..... 1419
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
18 rIeuAlaIeAlaIeugIyAlaIeugIyAlaIaProIaIaTyAla 35
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1420 ....GAAACCAAGCGTCGTCAGCGTGCGCGCAAGGTAAA..GTCATC 1461
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
35 sPTpAsnaAsnGlnSerIleIleIysAlaGlyGluArgGlnIaGlyIle 51
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1462 TTGATGATCGAGCGGCGCATCGCAAGCAAAACCAAGCCTTTATGAAAT 1511
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
52 HisIleYsIeuSerAspGlyAlaGlyValArgThrAlaThrGlyThr 68
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1512 CCGCTTCGTCAGCGGCGAGGCGAGCGTCACTGAATGCGATTAATCACT 1561
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 rIleIysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn.... 83
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1562 TCACCCCGACAACTATTCGCGCTTCGCGCGCGGAGCGCTTGATTTG 1611
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 .....ProIaIaGluLeuArgPheGlnAsnGlySerValThrSer 97
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1612 AACGGCATTCGCTTCGTCACCGCATCAAAATACCGATGAAGGCG 1661
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
98 SerGly.....GlnLeuPheAspIleGlyVal 106
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1662 GATGATGTCACCCACATCAAGACAAGAAATCCACGTTACATTACAG 1711
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
106 I.....ArgArgPheLeuGlyThrValThrValIysa 117
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1712 GCAATTA.....GATATTACTACACCGGCATATACCAACTTG 1752
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 IagIyIysLeuValAlaAspHisAlaThrLeuIaAsnValSerAspThr 133
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1753 GATGCAAAAAAGAAATTCG...TACACGTTGGTTTGCGAGAAAGA 1799
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
134 ArgAspAspAspGlyIleAlaIaLeuTyValAla....GlyIuGln.. 147
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1800 TGCACCAAAACGACGCGGCGCTCAATCTGAATTCACCAACGGAAG 1849
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
148 .....AlaGlnAlaSerIle 153
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1850 CGGATTCGACTTACTCTTCGCGCGGACACAAATTA.....AACGC 1893
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
153 IaaSPSerThrLeuGlnGlyAlaGlyValArgValGluArgGlyAla 169
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1894 AATATCAAGCAAAACGCGCAAACTGTTTTCAGCGCGAGCAGCACAC 1943
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 AsnValThrValGlnArgSerThrIleValAspGlyIysLeu..... 183
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1944 GCACGCTTACATCATTTAGGAAGCGGCTGTCA...AAATGAAAGTA 1990

```

```
184 .....HistIleGlyThrLeuGlnProLeuGlnProGlnAspL 196
1991 TCCACAGAGAGAAATCGTGTGGACAACGATTGGATCGACCGC..... 2034
196 euProProSerArgValValIleuGlyAspThrSerValThrAlaValPro 212
2035 .....ACATTAAAGCGGAAACCTCCA 2057
213 AlaSerGlyAlaProAlaIleValPheValPheGlyAlaAsnGlnLeuTh 229
2058 TMTTCAGGGCGGACAGCGGTGTTCCGCAATGTGCCAAAGTGAAG 2107
229 rValAspGlyGly..... 233
2108 GCGATTGGCATTTAAGCAATCAGCCCAAGCAGTTTGGTGGCGACCG 2157
234 .....HistIleThrGlyAlaArgAlaIle.....GlyValAla... 244
2158 CATCAAAAGCCACAACTGTACGCTTGAGCTGGACGGGTGTGACAAG 2207
244 ..... 244
2208 TTGTACCGAAAAACCATTTACGCGAGATTAAGTGTGCTTCATTGACCA 2257
245 .....AlaMetAspGlyAlaIleValHisLeuGlnA 255
2258 AGACCGACATCAGAGCAATGTACGCTTGCGGATCAGCGCTCATTAAT 2307
255 rGAlaThrIleArg..... 259
2308 CTCACAGGACTTGGCACACTCAGCGCAATCTTAGTCAGCGGAGACAC 2357
260 .....ArgGlyAspAlaProAlaGlyGly..... 267
2358 GCACATATAGCTTAGCGGACAGCGCCAAAGCGCAACCTC..... 2400
268 .....AlaValProGlyGlyAlaValProGlyGlyAlaValProGlyG 282
2401 .....AGC 2403
282 LypheGlyProLeuLeuAspGlyTyrGlyValAlaSpValSerAspSer 298
2404 CTCGTGGCGAATGCCCAAGCAACATTTAATCAAGCCATTAAGCGCA 2453
299 ThrValAspLeuAlaGlnSerIleValGlnAlaProGlnLeuGlyAlaAl 315
2454 CACATCGGCTTCGGCAATGCTTCAATTAACTAAGC..... 2490
315 AlIeArgAlaGlyArgGlyAlaArgValThrValSerGlyGlySerLeuS 332
2491 .....AACACGCGGTACAAACGCG..... 2511
332 erAlaProHisGlyAsnValIleGlnThrGlyGlyAlaArgArgPhe 348
2512 .....AGTGTAGCGCTTCCGACAAAGCTTAAGC 2540
349 ProProProAlaSerProLeuSerIleThrLeuGlnAlaGlyAlaArgAl 365
2541 AAGC.....GTAGCCCAT 2554
365 agInGlyArgAlaLeuLeuTyrArgValLeuProGlnProValIlyLeuT 382
2555 CCGGACTCAAGCGCAATGTCCTTCCCTAAGCGTAAGCGCAGTATTCATT 2604
382 hrLeuAlaGlyGlyAlaGlnGlyAsp.....IleValAlaThr 396
2605 GAAAAACAGCGCTTACCGGAAATACGCGGCGGCAAGATACGCAAT 2654
397 GluLeuProProIleProGlyAlaSerSerGlyProLeuAspValAlaIe 413
2655 ACACTTAAAGACAGCAATGAGCGTCCGTCGGGACAGCAATTAGCA 2704
413 u...AlaSerGlnAlaArgTyrThr...GlyAlaThrArgAlaValAspS 428
2705 ATTAAACCTTGACAGCCACCATTAACCTCACTCACTTCGCTATCGACAC 2754
428 erLeuSerIleAspAlaIleThrThrValIleMetThrAsp..... 440
2755 GATGCGGCGAGCGCGCAACCGGAGTGCAGAGATGCGCGCGCGCGCG 2804
441 .....AsnSerAsnValGlyAlaLeuArgLeuAl 450
2805 TTCGCGCGCTTCCCTATTAATCCGTTACGCGGCACTTCGCGAAATCC 2854
450 aserAspGlySer...ValasPheGlnGlnProAlaGlnAla...GlyA 465
2855 GTTTCACAGCGTGCAGCTTAACGCGCAATTAAGACGTCAGGACATTC 2904
465 rGpHeIysValIleuMetValAsp...ThrLeuAlaGlySerGlyLeuPhe 480
2905 CGCTTATGTGCGAACTCTTCCGCTACCGCGGCGCAATTAAGCTGCG 2954
481 ArgMetAsnValPheAlaAspLeuGlyLeuSerAspIlySerValIle 497
2955 GGAAGATTCGGAAGCGCACTTACACTTGGCTGTCAACATATACGCAAG 3004
497 IargAspAlaSerGlyGlnHisArgLeuTyrValArgAsnSerGlySerG 514
3005 AACCCGTAAGTCTGAGCAATGTACGCTAGTAGAGAGAAAGCAACACA 3054
514 IuproAlaSerGlyAsnThrMetLeuLeuValGln.....Thr 526
3055 CCGCTGTCCGAAATCTTAATTTACCTGCAAAAGCA.....CAGCT 3098
527 ProArgGlySerAlaIleThrPheThrLeuAlaAsnIlyAspIlySya 543
3099 CGATGCGCGCGCATGGCTTATCACTTATCCGAAA...GACGGGAGT 3145
543 IAspIleGlyTThrTyrArgTyrArgLeuAlaAlaAsnGlyAsnGlyGln 560
3146 TCCGCTGCATTAATCCGGTCAAGAAAGAGCTTCCGCAAACTCGCG 3195
560 rPserLeu.Val..... 563
3196 AAGCGGGAGAAACAGAGCGGCTTGACGCAAAACAGGCAACACTTGC 3245
563 ..... 563
3246 CCGCAAAACACAGCGGAAAAACACAGCGCAACGCTTGACGCGCTGA 3295
563 ..... 563
3296 TTGCGCGCGGCGCAATGCCACCGAAAGCAGAAAGTGTGCGCAACG 3345
564 .....GlyAlaIlyAsnIleProProAlaProIlyProAlaPro...Gl 576
3346 GCCCGGAGGAGCGGCGGAAAAATGCCGCAATTATGACGCGGAGAGA 3395
576 nProGlyProGlnProGlyProGlnProPro..... 586
3396 GAAAAAAGCGGTGCAGCGGATTAAGACACCGGCTTGCGGAAACAGCGG 3445
586 ..... 586
3446 AAGCGAAACCGCGCGGTACACAGCGCTTCCCGCGCGCGCGCGCG 3495
587 .....GlnProProGln.ProProGlnProProGlnP 597
3496 CGCGGAGATTTGCCGCAACCGACGCCCA.....CGCAACCCCA 3536
597 roPro.GlnProProGlnArgGlnProGlnAlaProAlaProGlnProPr 613
3537 ACCGCAAGCGGACACGATGACGCGTATGCAATAGCGGTTGAGTGAAT 3586
613 oAlaGlyArgGlnLeu...SerAlaAlaIleAsnAlaIleValAsnThrG 629
```

```

3587 TTTCGGCTGCTACAGCGTT.....TTCCGGCTAGAGAGATTG 3630
      ::::: ||| ||::: ::::: |||
629 IyglvalaIgluleuAlaserThrlleuTrpYrAlaIuSerAsnAlaleu 645
      ::::: ||| ||| |||
3631 GACCGCGTGTTCGGAAGACCGCCGAC.....GCCGTTCG 3668
      ::::: ||| ||| |||
646 SerIysArgIuIgluleuAlaIuSerProAspAlaIglYlAlar 662
      ::::: ||| ||| |||
3669 GACAAGCGGATCCGGACACCAACACTACCTTCGCAAGATTTCGCG 3718
      ::::: ||| ||| |||
662 pglYArgIglPhe.....A 667
      ::::: ||| ||| |||
3719 CTTACCGCAACAACCGACCTGCGCAATTCGTAAGAAAACTTC 3768
      ::::: ||| ||| |||
667 IaglnArglnglnIleuAsp..... 673
      ::::: ||| ||| |||
3769 GGCAGCGCGCGCTGGCATCTGTTCGCAACCGGACCGGAAACAC 3818
      ::::: ||| ||| |||
674 .....AsnArgIaIglYArg 679
      ::::: ||| ||| |||
3819 CTTCGAGAC.....GGCATCGCAACTCGGCAC 3847
      ::::: ||| ||| |||
679 gPheAspGlnIyValaIaIglPheGluIeuGlyAlaAspHisAlaVala 696
      ::::: ||| ||| |||
3848 GCGTTGCCACGCGTGC.....GTTTCGGCAATACGGCATCGGACG 3891
      ::::: ||| ||| |||
696 IavalAlaIglYArgTrpHisIeuIglYleuAlaIglYTrpHisArg 712
      ::::: ||| ||| |||
3892 TTGCATCGGATCAGCGCGCGCGGTTTATAGCGGACGCTT.. 3939
      ::::: ||| ||| |||
713 GlYAspArgIglPheTrpHisArgIglYglYHisIThrAspSerValH 729
      ::::: ||| ||| |||
3939 ..... 3939
      ::::: ||| ||| |||
729 sValIglYTrAlaThrlleAlaAsnSerGlyPheTrpIleuAspA 746
      ::::: ||| ||| |||
3940 .....TCA 3942
      ::::: ||| ||| |||
746 IatHrlleuArgAlaserArgIeuIuAsnAspPheIyValaIaIglYSer 762
      ::::: ||| ||| |||
3943 GACGCG.....ATCAGAGCAAAATCCGCCCGCGCTGCTGATTAAGG 3986
      ::::: ||| ||| |||
763 AspIglYrAlaValIySglYsTrYrArgTrHisIglYalIglYalSe 779
      ::::: ||| ||| |||
3987 CATTCAGGCAAGATACCGC..GCAGGTTTCGGGCGATTCGGCATCGAAC 4033
      ::::: ||| ||| |||
779 TrleuGlnAlaIglYArgArgPheAlaHisAlaAspGlyTrpPheIuSer 796
      ::::: ||| ||| |||
4034 CGCACATCGCGGCAACGCGCTATTCGTCGCAAAAGCGGATTAACGATAC 4083
      ::::: ||| ||| |||
796 roGlnAlaIgluleuAlaValaPheArgValaIglYglYalYalYArgAla 812
      ::::: ||| ||| |||
4084 GAAAC...GTCAATATCGCCACCGCGCTTCGATTCACCGGCTACCG 4130
      ::::: ||| ||| |||
813 AlaAsnGlnIleuArgValaIArgAspGlnIglYSerSerValIeuIgl 829
      ::::: ||| ||| |||
4131 CGCGGCGATTAAGCGAGATTAATTCATTCACACCGCGGCAACATTTCCA 4180
      ::::: ||| ||| |||
829 glleuGlyIeuIglValaIglYsArgIleGluIleuAlaIglYlArgIln 846
      ::::: ||| ||| |||
4181 TCACGCGTATTTAGCGCTGCTATTCAGATCGCGCTGCGCAAGATC 4230
      ::::: ||| ||| |||
846 alGlnProTrpYrIleuYalaserValIeuIuSerPheAspGlyAlaIgl 862
      ::::: ||| ||| |||
4231 CGAACGCGCTCAATACCGCGCTATTCGCGGAGATTCGCGCAAAACCG 4280
      ::::: ||| ||| |||
863 ThrValArgTrpHisArgIleAlaHisArgTrHisIleuArgIglYTrp 879
      ::::: ||| ||| |||
4281 CAGTSGGGAATGGGCGGTAACGCGGAAATCAAGTTTCACGCTGCGC 4330
      ::::: ||| ||| |||
879 g...AlaIgluleuIy.....IleuIy 886
      ::::: ||| ||| |||

```

```

4331 TCCAGCTGCCCGCCCAAGGGCCGCAATTGGAAGCC.....CAGCAC 4374
      :: ||||| ::::: ||| |||
886 eValAlaIaIaIleuIglYArgIglYHisSerIeuTrYAlaIaserTylglYr 902
      :: ||||| ::::: ||| |||
4375 AGCGCGGCATCAATTAAGCTACCGCTGG 4404
      :: ||||| ::::: ||| |||
903 SerIySglYProIySleuAlaIeuIleuProItrp 912
      :: ||||| ::::: ||| |||

```

```
seq_name: /cgn2_6/prodata/1/lae/5A_COMB.pep:US-08-038-682-4
```

```
seq_documentation_block:
```

```
; Sequence 4, Application US/08038682
; Patent No. 5549897
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: BARENKAMP, STEPHEN J
```

```
; APPLICANT: ST. GEME III, JOSEPH W
```

```
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
```

```
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
```

```
; NUMBER OF SEQUENCES: 8
```

```
; CORRESPONDENCE ADDRESS:
```

```
ADDRESS: Shoemaker and Mataré, Ltd
```

```
STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
```

```
STREET: Bldg. 1
```

```
CITY: Arlington
```

```
STATE: Virginia
```

```
COUNTRY: U.S.A.
```

```
ZIP: 22202-0286
```

```
COMPUTER READABLE FORM:
```

```
MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.25
```

```
CURRENT APPLICATION DATA:
```

```
APPLICATION NUMBER: US/08/038,682
```

```
FILING DATE: 16-MAR-1993
```

```
CLASSIFICATION: 424
```

```
ATTORNEY/AGENT INFORMATION:
```

```
NAME: BERSSTRESSER, JERRY W
```

```
REGISTRATION NUMBER: 22,651
```

```
REFERENCE/DOCKET NUMBER: 1038-293
```

```
TELEPHONE: (703) 415-0810
```

```
TELEFAX: (703) 415-0813
```

```
INFORMATION FOR SEQ ID NO: 4:
```

```
SEQUENCE CHARACTERISTICS:
```

```
LENGTH: 1477 amino acids
```

```
TYPE: amino acid
```

```
STRANDEDNESS: single
```

```
TOPOLOGY: linear
```

```
MOLECULE TYPE: protein
```

```
US-08-038-682-4
```

```
alignment_scores:
```

```
Quality: 274.50 Length: 1488
```

```
Ratio: 0.419 Gaps: 73
```

```
Percent Similarity: 44.019 Percent Identity: 19.220
```

```
alignment_block:
```

```
US-09-303-518D-653 x US-08-038-682-4 ..
```

```
Align seg 1/1 to: US-08-038-682-4 from: 1 to: 1477
```

```
178 AATAAGCAAGCTTTGCACTCGGGCGGAAAGATATTAGGTTTAAACAA 227
```

```
||| ||| ::::: ||||| ||||| ::::: |||
```

```
199 AsnHisGlyIleuIleThrlValaIgl...IyAspGlySerVal..... 211
```

```
228 AAAGGCGAGTGTGCGCAATCGATGACGAAGCCCGGATGATGATT 277
```

```
::: ||||| ::::: |||
```

```
212 .....AsnIleIleGlyIglYlYsValIyAsnIlu..... 221
```

```
278 TTTCTGTGTTATCGGCTAACGCG.....GTGGCGCATTTGGCGGCGAT 321
```

```
::: ||||| ::::: ||||| ::::: ||||| ::::: ||||| ::::: |||||
```

```

222 ..GlyValIleSerValAsnGlySerIleSerLeuLeuAlaGlyIn 237
322 CAA.....TATATTGCGAG 335
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIleAl 254
336 CGTGGCACAATAACGGCGCTATACAAATGTGATTTGGTCGGAGGAA 385
254 AlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaIleGlyG 271
386 GCAAT.....CCCGATCAGACCGCTTTTCT 411
271 LysAsnIleAsnValAlaGluAlaThrIleAlaGluAsnGlnGlyLeuSer 287
412 TACCAATTTGTGAAGAATAATTAATTAACGAGGACTACGGCCATCC 461
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTATGGCGCGATTAATCATATATGCGCGTTTGCACAAATTTGTCACAGATG 511
298 .....AsnIleValLeuSerA 303
512 CAGAACCTGTGAG.....ATGACGAGTTATATGAT 543
303 IalysGluGlyGluAlaGluIleGlyValIleSerAlaGluAsnGln 319
544 GGGTGAATACGCTGATTTAATAATATACCGTATGCTGT.....CG 587
320 GlnAlaIalysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLys 336
588 AATCGGAGCA.....GGCAGACAAATAT 610
336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyIleThrLys 353
611 GCGCGTCTGATGAA...GACGACCCCAATAC..... 639
353 euGlyLysAspGluArgGlyGlyLysAsnGlyIleGlnLeuAlaLys 369
640 .....CGGAAAGTTCAATCATATATGCAAGC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLys 386
667 .....GCATATCTCTTG.....CTCGTCGGTGGCATA 694
386 sGlyGlyPheAlaIleValIleArgGlyAspIleAlaLeuIleAspLysnI 403
695 CCTTGACAAATGANTGANTGAGTGTGGCAGCAGTCACTTAGTAGCGAA 744
403 IeAsnAlaGln...GlySerLys..... 409
745 AAAATTAACATAGCCCATATGTTTTCACACAGAGGCTCATTTGG 794
410 .....AspIleAlaLysThrGlyGlyPheValG 419
795 CGACAGTGGCTCACCAATGTTTATCTATGATCCCAAAAGCAAAAGTGT 844
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGTATGCAAAACAGCAACCCCTATATAGCAAAAGCAAT 894
430 .....AsnAlaIleValAspAla..... 435
895 GGCCTTCAGCTACTTCGTAAGATGTTCTATGATGAATCTTTGCTGG 944
436 .....LysGluThrLeuLeuAsp.....Ph 442
945 AGATACCATTCAGTATTTACGAA...CCACATCAAAATGGGAATACT 991
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGly 459
992 TTTTAACGACAATAATAATGCGCAGAAATAATGCATGCCAAACATATA 1041
459 IeAsnAspGluPheProThrGlyThrGlyAlaSerAspProLysLys 475
1042 CACTATCTCTACCTTATAGATTAACACAGACCGTCAATTTGTT... 1089
476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnIle 492
1090 .....AATGTTTCTTTATCCGAGACGACGACAGAAC 1120
492 uLysAsnAlaIlePheThrMetAsnIleThrAlaSerArgLysLeuThr... 507
1121 CTGTTTATCATGCTGCAGCTGGGCTCAACAGTTATCCAGCCAGACTGAAT 1170
508 .....ValAsnSer.....SerIleAsn 513
1171 AATGAGAAATATTTCTTTATTCACAAAGGAAAGGTAAATGTACT 1220
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1221 TACACGCAACATCAACCAAGCGCGCGCTTGTATTTTGAGGGTAAT 1270
527 .....GlnArgGlyGlyGlyValGlnIleAspGlyAspI 538
1271 TTACGCTTCGCTTAAAAACACGAAAGCTGGCAGGCGCGGCTTCA 1320
538 IeThrSerLysGlyGlyAsnLeuThrIleThrSerGlyGlyThrValAsp 554
1321 ATCAGTGAATGGCAGTACCGCTTACTTGAAAGTAACGGCGGTGCAACGA 1370
555 ValHisLysAsnIleThr..... 560
1371 CGGCTGTCCAAATGCGCAAGGACGCTGCTGTCAAGCC..... 1413
561 .....LeuAspGlnGlyPheLeuAsnIleThrAlaIleSerV 573
1414 .....AAAGGGAAACCAAGGCTCGTCAAGCTGGCGCAGCGTAA 1455
573 aAlaPheGluGlyGlyAsnAsnLysAlaArgAspAlaIleAsnAlaLys 589
1456 GTCATCTTGATCAGCAGCGGAGCATCAAGGCAAAACAGCCTTAG 1505
590 IleVal..... 591
1506 TGAATCGGCTTGTCAGCGGACGCGGAGCTGCACTGAATGCCGATA 1555
592 .....AlaGlnGlyThrValThrIleThrGlyGluG 602
1556 ATCAGTTCAACCCGACCAACTATTTGCGCTTCGGCGCGGAGCTTGG 1605
602 Lyls.....AspPheArgAlaAsnAsnVal 610
1606 GATTGACGGGCAATTCGCTTTCGTTCCACCGCATTAATAATACCGATGA 1655
611 SerLeuAsnGly.....ThrGly 617
1656 AGGGCGGATGATTTGC.....AACCAATCAACGACAAAGANT 1693
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACGCTTACATTCAGGCAATTAAGATATTAAGTACACCGGCAATAC 1743
634 LysThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerLysThrPheIleThrSerHisAspSerHisThrAsnValSerAl 667
1747 ....AACTTGAT..... 1755
667 aLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyrlIleSers 684
1756 .....AGCAAAAGCAATTCGC 1773
684 eAsnSerLysGlyLeuThrThrGlnTyArgSerSerAlaGlyValAsn 700

```



```

1774 TACACGGTGGTGGTGGGAGAAAGATGCAACCAAAAGAACGGCGCT 1823
      :|||:|
      :|||:|
701 pheasnly.....ValasnlyAsnme 708
1824 CAATCTGATATACCAACGGAGAGCGGATGCACCTTTACTGTTCCG 1873
      :|||:|
      :|||:|
708 tSerPheasnleuys...GluGlyAlaLysValAsnPheLysleuysp 724
1874 GCGGACAAATTTAAAC.....GGC 1893
      :|||:|
      :|||:|
724 roasnGluAsnMetasnThrSerIlyProleuprolleatpHeleuAla 740
1894 AATATCAGCAACA...AACGGCAACTGTTTTCAGCGGAGACCGAC 1940
      :|||:|
      :|||:|
741 AsnIleThrIleThrGlyGlySerValPhePheasp..... 753
1941 ACCGCGCGCTAC...AATCATTTAGAACGGGCTGTCAAAATGAAG 1987
      :|||:|
      :|||:|
754 .....IleThrAlaAsnHisSerGlyArgly.....AlaG 764
1988 GTATCCCAAGAGAAATCGTGTGGACAACGATTGATGACCGCACA 2037
      :|||:|
      :|||:|
764 IuleuLysMetSerGluIleAsnIleSerAsn..... 774
2038 TTTTAAACGGAAACTTC.....CATATTCAGGCGGAGACAGC 2075
      :|||:|
      :|||:|
775 .....GlyAlaAsnPheThrLeuAsnSerHisValArgIlyAspAspAl 789
2076 GGTGGTTCCCGCAATGTGCCAAAGTGAAGGCGATGGCATTTAAGCA 2125
      :|||:|
      :|||:|
789 aspHe.....LysIleAsnLysAspLeuThrIleAsn. 799
2126 ATCAGCCCAAGCAGATTTCGTGTGCGACCGCATCAAGCCACACATC 2175
799 ..... 799
2176 TGTACAGCTTCGACGTGAGCGGTGTGACAGATTGACCAAAACCAT 2225
      :|||:|
      :|||:|
800 AlathrAsnSerAsnPhe..... 805
2226 TACCGAGATAAAGTATGCTTCATTTGACAGACCGACATCAGAGCA 2275
      :|||:|
      :|||:|
806 .....SerLeuArgGlnThrLysAspAspPheT 815
2276 ATGTACGCTTGGCGATCAGCTCATTTAATCTCACAGCATTCGCACA 2325
      :|||:|
      :|||:|
815 yIaspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 831
2326 CTCAGCGCAATCTTACTGACGCGGAGACACGCACTATACGGTTACGCG 2375
      :|||:|
      :|||:|
832 LeuGlyGlyAsnValThrLeuGlyGly..... 840
2376 CAACGCCAACCAACGGCAACCTCAGCCTCGTGCGAATGCCCAACGA 2425
      :|||:|
      :|||:|
841 .....GlnAsnSerSerSerIleThrGlyAsn.....IleT 852
2426 CATTTAATCAAGC.....ACATTAACGGCAACACATCGGCTCG 2466
      :|||:|
      :|||:|
852 hrIleGlyLysAlaIleAsnValThrLeuGluAlaAsnAlaProAsn 868
2467 GACAAATGCTTCATTTAATCTAAGCAACAACGCGGTACAAAACGGCAGTCT 2516
      :|||:|
      :|||:|
869 GlnGln.....AsnIleArgAspArgValIleLysLeuGlySerIle 882
2517 GAGCGTTTCGACACAGCTAAGCAACGTAAGCCATTCGCGACCTGCAAGC 2566
      :|||:|
      :|||:|
882 uLeuValAsnGlySerLeuSerLeuThrGlyLysAlaAspIleLysG 899
2567 GCAATGTCTCCAGCGGATAGCAGATATTCAT.....TTTGA 2607
      :|||:|
      :|||:|
899 LysnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCGCGTTTACGCGAAATACAGCGCGGAGCAAGATACGCAATTACA 2657

```

```

      :|||:|
      :|||:|
916 ThrLeuAsnIleThrGlyLysnPherThrAsnAsnGlyThrAlaGluIleAs 932
2658 CTTAAAGACAGCAATGAGACCTCGCTGGGCAACGGAATTAGCAATT 2707
      :|||:|
      :|||:|
932 nIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspL 949
2708 TAAACCTTGACAAGCGCACCATTTACATCTCATTTCCGCTATGACACGAT 2757
      :|||:|
      :|||:|
949 euAsn.....IleThrThrHisAla..... 955
2758 GCGGAGCGCGCAAAACCGCAGTCGCGAGATGCGCGCGCGCGCTTC 2807
      :|||:|
      :|||:|
956 .....LysArgAs 958
2808 GCGCGCTTCCTATTATCCGTACCGCGCAACTTCGCGACATCCGCTT 2857
      :|||:|
      :|||:|
958 ngIleArgSerIle..... 962
2858 TCACACGCTGACGGTAAACGCAATTTGAACGTCGAGGAAACATTCCGC 2907
      :|||:|
      :|||:|
963 .....IleGlyGlyAspIleIleAsn..... 969
2908 TTTATGTCGAACTTTCGCTACCGCAGCGCAATTTGAACCTGCGCA 2957
      :|||:|
      :|||:|
970 .....LysLysGlySerLeuAsnIleThrAs 978
2958 AAGTTC.....G 2965
      :|||:|
      :|||:|
978 pSerHisAsnAspAlaGluIleGlnIleGlyLysnIleSerGlnLysG 995
2966 AAGGCACTTACACTTGCT.....GTCAACATACCGCAACGAA 3006
      :|||:|
      :|||:|
995 IuGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr..... 1008
3007 CCGGTAGTCTCGACCAATTTGACGTA.....GTGGAAGAAAGA 3047
      :|||:|
      :|||:|
1009 .....LysGlnIleThrIleLysLysGlyIleAspGlyLys 1021
3048 C...AACACACGCTGTCCGAAATCTTAATTTCCCTGCAAAACGAC 3094
      :|||:|
      :|||:|
1021 pSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIleLys..... 1035
3095 ACGTCGATCCGCGCATGCGCTTATCAGCTATCCGAAAGACGCGAG 3144
      :|||:|
      :|||:|
1035 ..... 1035
3145 TTCGCGCTGCATTAATCCGCTCAAGACAGAGCTTTCGACAAACTCG 3194
      :|||:|
      :|||:|
1036 .....ThrLysGluLeuLysLeuThrGluAspLeuSe 1046
3195 CAAGCGGAGAAACAGAGCGCGCTTACGCGCAAAACAGCACAACCTG 3244
      :|||:|
      :|||:|
1046 rIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAspL 1063
3245 CCGCAACACACAGCGGAGAAAGACAAACGCGCAAGC..... 3282
      :|||:|
      :|||:|
1063 euThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLysThr 1079
3283 .....CTTGACGCGCTGATTGCGGC...GGCGGCA 3311
      :|||:|
      :|||:|
1080 ValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHisAs 1096
3312 TGGCACCGCAAAAGCAGAAAGTGTTCGCGCAACGCGCGCGGAGCGAG 3361
      :|||:|
      :|||:|
1096 nValThr.....LeuAsnSerLysValLysThrSerSerSerAsnGlyG 1111
3362 GCGAAATGCGCGCATTTATGACGCGGAGAGAAAGAAACAGGTCGAG 3411
      :|||:|
      :|||:|
1111 LysArgGluSer.....Asn 1115
3412 GCGGATTAAGACACCGCTTG.....CGGAACAGCGCAAGCGCA 3452
      :|||:|
      :|||:|

```

```

1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal.. 1131
3453 AACCCGGGGGCTACCAACGGCTTCCCGGGGGCCGGCCGGCCGGG 3502
1131 ..... 1131
3503 ATTTCGGCAACCGCAGCCCAACCGCAACCCCAACCGCAGCGGACTG 3552
1132 ..... 1132
3553 ATCAGC...CGTATGCCATAGCGGTTTGAGTGAATTTTCCGCGACGCT 3599
1133 ..... 1133
1136 ThrSerLeuYsthrValAsnIleThrAlaSerGlyLysValThrThr..T 1152
3600 CAACAGCGTTTTCGCGTACAGAGCAATTCGAGCGGCTGTTTCCGCAAG 3649
1152 hralaglySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168
3650 ACCCGCGCAACGCGGTTTGCAACAGCGGCAATCCGGAGACCAACACTAC 3699
1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185
3700 CGTTCGC...AAGATTCCGGCGCTACCGCCCAACAAACCCACCTCGCCCA 3746
1185 rAlatThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysS 1202
3747 AATCGGATTCGAGAAACCTCGGAGCGGCGGCGGCTGCTGTTT 3796
1202 erglyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly..... 1216
3797 CGCACACCGGACCGGAAACCTTCGACGACGAGCATCGCA..... 3838
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232
3839 ..... 1232
1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrL 1249
3861 TGCCGTTTTCGGCAAT.....ACGGCATCGGAGGTTGCACATTCGCA 3904
1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265
3905 TCAGCGCGCGCGGCTTTAGTAGCGGACGCTTCAGACGCGCATGAGA 3954
1266 SerThrLysGlyGlnValAspLeuIleAla.....GlnAsnGlySerI 1280
3955 GGCAAAATCGCGCGCGGCTGTCATTCAGCATTCAGCGCAATATCCG 4004
1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297
4005 CGCAGGTTTCGGCGGATTCGCGATCGCAACCGCACATCGCGCAACGGCT 4054
1297 euThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313
4055 ATTTCGTCGCAAA 4066
1314 IleAsnAlaLys 1317

```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-302-832-4

```

seq_documentation_block:
; Sequence 4, Application US/08302832
; Patent No. 5603938
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: Shoemaker and Mathare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington

```

```

STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,832
FILING DATE: 16-SEP-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US pct/us93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Belkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-302-832-4

```

```

alignment_scores:
Quality: 274.50 Length: 1488
Ratio: 0.419 Gaps: 73
Percent Similarity: 44.019 Percent Identity: 19.220

```

alignment_block:
US-09-303-518d-653 x US-08-302-832-4 ..

```

Align seg 1/1 to: US-08-302-832-4 from: 1 to: 1477
178 AATAAGGCAAGTTTCAGTTCGGGCGGCAAGATATGAGTTTACACAA 227
199 AsnIleGlyLeuIleThrValGly...LysAspGlySerVal..... 211
228 AAAAGGGAGTTGGTGGCAATTCGATGACGAAAGCCCGATGATGATT 277
212 .....AsnLeuIleGlyGlyLysValLysAsnGlu..... 221
278 TTTCGTGGATCGCTAACGGC.....GTGGCGCATTCGCGGCGCAT 321
222 ..GlyValIleSerValAsnGlyLysSerIleSerLeuLeuAlaGlyGln 237
322 CAA.....TATATTGTGAG 335
238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleL 254
336 CGTGCACATACGCGCGCTATACATGTTGATTTGGTCGGAGGCA 385
254 aAlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGly 271
386 GCAT.....CCGATTCAGCAGCGCTTTCT 411
271 LysAsnIleAsnValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSer 287
412 TACCAATTCGAAAGAAATATATTAAAGCAGGAGTAAGCGCATCC 461
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTATGGCGGATTAACATATATGCGCGGTTTGCAAAATTTGTCAGATG 511

```

298:|||||:..:|
.....AsnIleValIleuSerA 303
512 CAGAACCTGTTAG.....ATGACCACTTAATATGAT 543
|||||:|
303 IalysGIuGIyGIuAlaGluIleGIyGIyAlIleSerAlaGlnAsnGln 319
544 GGGTGAATACGCTGATTAAATATACCTGATCGTGT.....CG 587
|||:|
320 GlnIalalysGIyGIyLysIleuMetIleThGIyAspIysValIthrIleuLy 336
588 AATGGAGCA.....GGCAGACAAATATT 610
:|||||
336 sThGIyAlaValIleAspIeuSerGIyLysGIuGIyGIuThrIyTL 353
611 GCGCGTGTGATGAA...GACGAACCCCAATAC..... 639
:|||||
353 euGIyGIyAspGIuArgGIyGIuGIyLysAsnGIyIleGlnIleuAlaLys 369
640CCGAAAGTTCATATCATATTCAGAC..... 666
370 LysThrSerIeuGIuLySGIySerThrIleAsnValSerGIyLysGIuLy 386
667GCATATTCCTTG.....CTCGTGGTGGCAATA 694
|||:|
386 sGIyGIyPheAlaIleValItrpGIyAspIleAlaIleuIleAspIylAsnI 403
695 CCTTGCACAAATGATGCAGGTGGTGGCACAGTCAACTAGATGCGAA 744
|||||:|
403 IeAsnAlaGln...GIySerGIy..... 409
745 AAATTTAAACATAGCCCATATGTTTTTTTACAAAGAGGCTCATTTGG 794
410AspIleAlaLysThrGIyGIyPheValGI 419
795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCAAAGCAAAAGTGT 844
|||||:|
419 uThrSerGIyHisAspIeuPheIleLysAsp..... 429
845 TAAATATGGGTATGCAACAGCAAGCAACCCCTATATAGAAAAAGCAAT 894
|||||:|
430AsnAlaIleValAspAla..... 435
895 GCGTTCAGCTAGTTGCTAAAGATTGGTCTATGATGAATCTTGTGG 944
|||||:|
436LysGIuTTrpIleuAsp.....Ph 442
945 AGATACCCATTCACTATTTACGAA...CCACATCAAAATGGGAAATACT 991
|||||:|
442 eAspAsnValSerIleAsnAlaGluAspProIeuPheAsnAsnThrGIyI 459
992 TTTTAAAGCAATATATAGCGCGCAAGAAATCGATGCCCAACATATA 1041
:|||||:|
459 IeAsnAspGIuPheProIthrGIyThrGIyAlaIleSerAspProLys 475
1042 CACTATCTCTACCTTATAGATTAAAGCAAGCAACCTCAATGTGT.. 1089
:|||||:|
476 AsnSerGIuIleuLysThrIthrIleuThrAsnThrIleSerAsnIyTL 492
1090AATGTTCTTTATCCGAGACAGCAAGAGAAC 1120
|||||:|
492 uLysAsnAlaItrpIthrMetAsnIleThrAlaSerArgLysIleuThr... 507
1121 CTGTTTATCATGCTGCAAGTGGGTCACAGTTATGCAACCCAGACTAAT 1170
508ValAsnSer.....SerIleAsn 513
1171 AATGAGAAATATATCTTTATGACAAAGGAAAGGATGATGATACT 1220
|||||:|
514 IleGIySerAsnSerIleuIleuIleuIleuIleuIleuIleuIleuIleu 526
1221 TACCAAGCAATCAACCAAGCGCGGCTTGTATTTGAGGGTAAAT 1270
|||:|
527GlnArgGIyGIyGIyAlaGlnIleAspGIyAspI 538
1271 TTACGGTCTCGGCTAAACCAACGAACGTGGCAAGCGCGGCGCTCAT 1320
:|||||:|
538 IeThrSerIysGIyGIyAsnIleuThrIleTySerGIyGIyTrpValAsp 554
1321 ATCAGTGAATGCGAGTACCGTTACTTGAAGAATAAACGCGCTGCAACAGA 1370
:|||||:|
555 ValHisLysAsnIleThr..... 560
1371 CCGCTGTCCAAATCGCAAGGACGCTGCTGTTCAAGC..... 1413
561LeuAspGIuGIyPheIleuAsnIleThrAlaIleSerV 573
1414AAGGGAAGAAACCAAGGCTGCTGAGGCTGGCGCAACGTAA 1455
:|||||:|
573 AlaIaPheGIuGIyGIyAsnAsnLysAlaArgAspAlaIleAsnAlaLys 589
1456 GTCATCTTATGATCAGCAGCGGAGCATCAAGCAAAAAACAAGCCTTAG 1505
:|||||:|
590 IleVal..... 591
1506 TGAATCGGCTTGTCAGCGGAGCGGAGCGGTGCACTGAATGCCGATA 1555
592AlaGlnGIyThrValIthrIleThrGIyGIuG 602
1556 ATCAGTTCAACCCGACAAACTATTTGGCTTTCGGCGGCGGAGCTTGG 1605
:|||||:|
602 Lys.....AspPheArgAlaAsnAsnVal 610
1606 GATTTGACGGGCATTCGCTTGTTCACCCCATTTCAAAATPACCAATGA 1655
:|||||:|
611 SerIleAsnGIy.....ThrGIyLy 617
1656 AGGCGGATGATGTTG.....AACCAATCAAGCAAAAGAT 1693
617 sGIyIleuAsnIleIleSerSerValAsnAsnIleuThrHisAsnIleuSerG 634
1694 CCACGCTTACCATTACAGCATTAAGATATTACTACAAACCGGCAATAC 1743
:|||||:|
634 LyrThrIleAsnIleSerGIyAsnIleThrIleAsnGIuThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerTyTrpIleThrIleThrSerHisAspSerHisTrpAsnValSerAl 667
1747AAGTTGAT..... 1755
667 AleuAsnIleuGIuThrGIyAlaAsnPheThrPheIleLysTyIleSers 684
1756ACGAAAAAAGAATTGCC 1773
684 eAsnSerIysGIyIleuThrIthrGlnTyArgSerSerAlaGIyAlaAsn 700
1774 TACAAAGTGTGTTGGCGAGAAAGATGCAACAAAGCAAGCGCGCT 1823
:|||||:|
701 PheAsnGIy.....ValAsnGIyAsnIle 708
1824 CAATCTGAATTACCAACCGGAAGACGAGTGCACCTTACTGCTTTCG 1873
:|||||:|
708 tSerPheAsnIleuLys...GIuGIyAlaIleValAsnPheLysIleuLysP 724
1874 GCGGAACAATTTAAAC.....GGC 1893
724 roAsnGIuAsnMetAsnThrSerLysProIeuProIleArgPheIleuAla 740
1894 AATATCAGCAAAACA...AACGCAAACTGTTTACAGCGGACGACGAC 1940
:|||||:|
741 AsnIleThrAlaIleThrGIyGIyLysSerValIlePheIleAsp..... 753
1941 ACCGACGCGCTAC...ATCATTTAGGAAGCGGCTGTCAAAATATGGAAG 1987
:|||||:|
754IleTyAlaIleAsnIleSerGIyArgGIy.....AlaG 764


```

1169 LysThrGlyAspIleSerGlyThrIleSerCylAsnThrValSerIse 1185
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
3700 CGTTGCC...AGATTTCGGCGCCTACCCGCCAACAGCACTGGCGCA 3746
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1185 rIatnThrValAspLeuThrThrIysSerGlySerIleGluIaIySs 1202
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3747 AATCGGTATGACAGAAAAACCTGGCGACGGGGCGCGTGCATCTGTTT 3796
      ::::: ||| ||| ||| ||| ||| ||| ||| |||
1202 erGlyIuIaIasnValIThrSerIatnThrIleGlyGly..... 1216
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3797 CGCACACCGGAGCCGAAACACTTGGAGACGGCATCGGCA..... 3838
      ||| ::::: ::::: ||| ::::: ||| ||| ||| ||| |||
1217 ...ThrIleSerGlyAsnThrValAsnValIThrAlaIasnIaGlyAspIe 1232
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3839 .....ACTCGGACGGCTTGGCCACGG 3860
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
1232 uThrValGlyAsnGlyAlaIaGluIleAsnIatnThrIuGlyAlaIaIhTrL 1249
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
3861 TGCGCGTTTCGGGCAAT....ACGGCATCGGCGAGGTTCGACATCGGCA 3904
      ::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1249 euThrIatnThrGlyAsnThrLeuThrIthrGluIaGlySerIleThr 1265
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3905 TCAGCCCGGGCGCGGCTTTTAGTACGGCGCACGCTTTAGACGCATCAGA 3954
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1266 SerThrIyGlyGlyIaIaIAspLeuLeuIaIa.....GlnAsnGlySerI 1280
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3955 GCGAAATATCCGCCCGCGCGTCTGCACNTACCGCATTTAGGCAAGATACCG 4004
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1280 eAtnIaGlySerIleAsnAlaIaIaIAsnValIThrLeuAsnThrThrGlyThrL 1297
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4005 CGCAGGTTTCGGGGATTCGCATCGAACCGCATCGGGCGACAGCGCT 4054
      ::::: ||| ||| ||| ::::: ||| ||| ||| :::::
1297 euThrThrValIaIaGlySerAspIleIySsIatnThrSerGlyThrLeuVal 1313
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4055 ATTTCTGCTCCAAA 4066
      ||| ::::: ||| ||| ::::: ||| ||| ::::: |||
1314 IleAsnAlaIySs 1317
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
seq_name: /cgn2_6/ptodata/1/1aa/5b_COMB.pep:US-08-530-198-4

```

667GCATATCTGTG.....CTGCTGGTGCATA 694
386 sglYgIyPheAlleValItrPglyAspIleAlaLeuIleAspIlysnI 403
695 CTTTGCACAAATGGATCAGTGGTGGCAGTCACTTAGTAGCGAA 744
403 IeAsnAlaGln...GlySerGly..... 409
745 AAAATTAACATAGCCCATATGTTTTTTTACCAACAGAGGCTCATTTGG 794
410AspIleAlaLysThrGlyGlyPheValGI 419
795 CGACAGTGGCTCCACCATGTTTATCTATGATGCCCAAAAGCAATGGT 844
419 uThSerGlyhIAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGTATTGGCAACAGCGCAACCCCTATATAGCAAAAGCAAT 894
430AsnAlaIleValAspAla..... 435
895 GGCCTTCAGTAGTTCGTAAAGATTGGTTCTATGATGAATCTTTGCTGG 944
436LysGluTrpLeuLysP.....Ph 442
945 AGATACCATTCAGTATTCTACGAA...CCACATCAAAATGGGAATACT 991
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnThrGlyI 459
992 TTTTAAAGCAATATATATGGCGGCAAAATCGATGCCAAACATATAA 1041
459 IeAsnAspGluPheProThrGlyThrGlyAlaIleAspProLysLys 475
1042 CACTATTCCTACCTTATAGATTAAACACAGAACCTTCATTTGTTT.. 1089
476 AsnSerGluLeuLysThrThrLeuThrAsnThrIleSerAsnTyIle 492
1090AATGTTCTTATCCGACACAGCAAGACAGAAC 1120
492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr... 507
1121 CTGTTTATCATGCTGCAGGTGGTGCAGGTTATCGACCCAGACGTGAT 1170
508ValAsnSer.....SerIleAsn 513
1171 AATGGAGAAATATTTCTTATATGACAAAGAAAGTGAATGATACT 1220
514 ILeGlySerAsnSerIleLeuIleLeuHisSerGly..... 526
1221 TACCAAGCAATCTCAACCAAGCGCGGCTTTGATTTAGAGGTAATT 1270
527GlnArgGlyGlyValGlnIleAspGlyAspI 538
1271 TTACGGTCTGCGCTAAACAAACAGCAAGCGGCGGCGGCTCAT 1320
538 IeThrSerIyGlyGlyAsnLeuThrIleTySerGlyGlyTyTrpValAsp 554
1321 ATCAAGTATGCGCAGTACCGTTACTTGGAAATGAACGCGCTGGCAACGA 1370
555 ValHisLysAsnIleThr..... 560
1371 CCGCGTGTCCAAATCGCAAGACGCTGCTGTCAAGCC..... 1413
561LeuAspGlnGlyPheLeuAsnIleThrAlaIleAsnServ 573
1414AAAGGGAAACCAACAGCTCGGTGACGCGTGGCGAGGTA 1455
573 AlaIlePheGlnGlyGlyAsnAsnLysAlaArgAspAlaIleAsnAlaLys 589
1456 GTCATCTTAATCATGACGCGGAGCATCAAGCAAAACAGCCTTTAG 1505
590 ILeVal..... 591
1506 TGAATATGGCTGTGTCAGCGGAGGCGACGTGCAACTGAATGCCGATA 1555
592AlaGlnGlyThrValThrIleThrGlyGlnG 602
1556 ATCAGTTCAACCCCGACAAACTATTTGCGGCTTGCAGCGAGCTTG 1605
602 Lys.....AspPheArgAlaAsnAsnVal 610
1606 GATTTGAACGGGCATTCGCTTTCGTCACCGCATTTCAAAATACCGATGA 1655
611 SerLeuAsnGly.....ThrGly 617
1656 AGCGCGATGATGTC.....AACCAATCAAGCAAAAGAT 1693
617 sglYLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACGCTTACCATTTACAGCAATTAAGATATTACTACCAACGCGCAATAC 1743
634 LThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerTyTrpIleThrIleThrSerHisAspSerHisTrpAsnValSerAl 667
1747AACTGGAT..... 1755
667 AleuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyIleSerS 684
1756AGCAAAACAAATTCGC 1773
684 eAsnSerLysGlyLeuThrThrGlnTyArgSerSerAlaGlyValAsn 700
1774 TACACGGTTGGTTGGCGAGAAAGATGCACCAACCAACAGACGCGGCT 1823
701 PheAsnGly.....ValAsnGlyAsnMe 708
1824 CAATCTGAATTTACCAACCGGAAGCGGATCGCACTTACTGCTTTCG 1873
708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724
1874 GCGGACCAAAATTAAC.....GGC 1893
724 roIsnGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATCCGCAACAAACA...AACGCAAACTGTTTTCAGCGGCGACCGAC 1940
741 AsnIleThrAlaThrGlyGlyIleSerValAlPhePheAsp..... 753
1941 ACCGCAAGCTTAC...AATCATTTAGGAAGCGGTGTCATAAATGGAAG 1987
754IleTyAlaAsnHisSerGlyArgGly.....AlaG 764
1988 GTATCCCAACAGAGAAATCGTGGGACACAGCATTTGATCGACCGGACA 2037
764 IuLeuLysMetSerGluIleAsnIleSerAsn..... 774
2038 TTTAAAGCGAAACTTC.....CATATTCAGGGCGGACGAC 2075
775GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGGTTTCCGCAATGTTGCCAAAGTGAAGCGCATTTGATTAAACA 2125
789 aPhe.....LysIleAsnLysAspLeuThrIleAsn. 799
2126 ATCAGCGCCACAGCATTTTTCGCTGTCGACCGCATCAAGCACACATTC 2175
799 799
2176 TGTACAGTTCGACTGAGCGGCTGTGACAACTGTACGAAAAACCAT 2225
800 AlaThrAsnSerAsnPhe..... 805
2226 TACCGACATTAAGATGTTGCTTCATTGACAGACCGACATCAGAGCA 2275

806SerLeuArgInThrLysAspPheT 815
2276 ATGTCAGCGCTTGGCATCAGCGCTTAATATCCTACAGGACCTTGGCACA 2225
815 yraspglytyrAlaArgAsnAlaIleAsnSerThrTyraSnIleSerIle 831
2326 CTCACAGCGCAATCTTATGCGAGCGGAGACAGCAGCTATACGGTTACGCG 2375
832 LeuGlyLysnValThrLeuGlyGly..... 840
2376 CAACGCCACCCAAACGCGCACTCAGCGCTGCGCAATGCCAACGAA 2425
841GlnAsnSerSerSerIleThrGlyAsn.....IleT 852
2426 CATTTAATCAAGCC.....ACATTAAAGCGCAACACATCGGCTTCG 2466
852 hrIleGlyLysAlaIleAsnValThrLeuGlnAlaAsnAlaProAsn 868
2467 GACAATGCTTCATTATTAATCTAAGCAGACAGCGCTACAAACGCGAGCTT 2516
869 GlnGln.....AsnIleArgAspArgValIleLysLeuGlySerIle 882
2517 GACGCTTCCGACAGCGCTAAGCGCAACGTAAGCCATTCCGACACTCAGC 2566
882 uLeuValAsnGlySerLeuSerLeuThrGlyLysAlaAsnAlaAspIleLysG 899
2567 GCAATGTCCTCCCTAGCGCATAGCAGTATTCAT.....TTTGAA 2607
899 LysnLeuThrIleSerGlnSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCGCGCTTACCGGCAAAATACAGCGCGGAGATACGGCATTTACA 2657
916 ThrLeuAsnIleThrGlyAsnPheThrAsnGlyThrAlaGluIleAs 932
2658 CTTAAAGACAGCGAATGAGCGCTGCGCGGAGCGAATTAAGCATTT 2707
932 nIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspIleAspL 949
2708 TAAACCTTGACACGCGCACCTTACACTCAATTCGCGCTATTCGACAGAT 2757
949 euAsn.....IleThrThrAla..... 955
2758 GCGGAGCGCGCAACCGGCGAGTGGCGAGATGCGCGCGCGCTTC 2807
956LysArgAs 958
2808 GCGCGCTTCCCTATATTCGTTAGCGCGCCACACTTCGGCAAAATCCGCTT 2857
958 nGlnArgSerIle..... 962
2858 TCACACGCTGACGTAACGCGCAATTTGAACGGTCAAGGAGACATTCGCG 2907
963IleGlyLysPheIleAsn..... 969
2908 TTTATGTCGAACCTTTCGGCTACCGGAGCGGCAATTTGAAGTGGCGGA 2957
970LysLysGlySerLeuAsnIleThrAs 978
2958 AAGTTC.....G 2965
978 pSerAsnAsnAspAlaGluIleGlnIleGlyLysnIleSerGlnLysG 995
2966 AAGCAGCTTACACCTTGCT.....GTCAACATTAACGGGCAACGAA 3006
995 LngLysnLeuThrIleSerSerAspLysIleAsnIleThr..... 1008
3007 CCCGTAAGTCTGAGCAATTCAGCGTA.....GTGGAAGGAAAGA 3047
1009LysGlnIleThrIleLysLysGlyLysAlaAspGlyLys 1021
3048 C..AACACACCGCTGTCGCAAAATTTAATTTCACCTGCAGAAACGAAAC 3094
1021 pSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIleLys..... 1035

3095 ACGTCAGTCGCGCGCATGCGCTTATCAGCTTATCCGCAAAAGACGCGGAG 3144
1035 1035
3145 TTCGCGCTGCATTAATCCGGTCAAGAACAGAGACTTTCGCAAACTCGG 3194
1036ThrLysGlnLeuLysLeuThrGlnAspLeuSe 1046
3195 CAAGCGGAGAAACAGACAGCGCCCTTACAGCGCAAAACAGCAGCAACTTG 3244
1046 rIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAspL 1063
3245 CCGCCAAACACAGCGCGGCAAAAGAACACAGCGCAAGC..... 3282
1063 euThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaIleAlaLysThr 1079
3283CTGACGCGCTGATTGGCGCC...GGCGGCA 3311
1080 ValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHisAs 1096
3312 TGCACCGCAAAAGCGAAGAGTGTTCGCAACCGCGCGCGGAGCGAGCG 3361
1096 nValThr.....LeuAsnSerLysValLysThrSerSerSerAsnGlyG 1111
3362 GGGAAATGCGCGCATTTATGAGCGGAGAGAGAAACGCGTGCAG 3411
1111 LysArgLysSer.....Asn 1115
3412 GCGGATTAAGACACCGCGCTTG.....GCGAAACACGCGAAGCGGA 3452
1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal.. 1131
3453 AACCGCGCGGCTACACCGCGCTTCCCGCGCGCGCGCGCGCGCGG 3502
1131 1131
3503 ATTTCGCGAACCGGACGCCCAACCGCAACCCCAACCGGAGCGGACCTG 3552
1132AsnLysAspIle 1135
3553 ATCAGC...CGTTATGCCAATAGCGGTTTGAATGTAATTTCCGCGACGCT 3599
1136 ThrSerLysThrValAsnIleThrAlaSerGlnLysValThrThr..T 1152
3600 CAACAGCGTTTTCGCGCTACAGAGCAATTSAGCCGCTGTTCGCGAAG 3649
1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168
3650 ACCGCGCAACGCGCTTTCGACAAGCGGATCCGGAGACCAACACATAC 3699
1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185
3700 CGTTGCG...AGATTTCGCGCGCTTACGCGCAACAAACGAGCTGCGCA 3746
1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLys 1202
3747 AATCGGTATGCAAGAAACCTTCGCGAGCGCGCGCTTCGCGATCTGTTT 3796
1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyLys 1216
3797 CGCACAAACGCGGCAACACCTTCGACGAGCGGATCGGCA..... 3838
1217 ..ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232
3839ACTGCGACAGCGCTTGGCCACG 3860
1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGlnGlyAlaAlaThrL 1249
3861 TCGCGTTTTCGCGCAAT.....ACGGCATCGCGAGTTTCGACATCGGCA 3904
1249 euThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265


```

845 TAATTATGGGCTATTGCAAAACAGGCAACCCCTATATAGAAAAAGCAAT 894
    |||:::|||||:::
430 .....AsnAlaIleValAspAla..... 435
895 GGCTTCCAGTAGTTCGTAAGATTGGTTCATGATGAATCTTGGTCGG 944
    |||:::|||||:::
436 .....LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCATTCAGTATTTCTACGAA...CCACATCAAAATGGCAATACT 991
    |||:::|||||:::
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459
992 TTTTAAACGACAAATATATAGCCGAGAAAATCGATGCCAAACATATA 1041
    |||:::|||||:::
459 IeAsnAspGluPheProThrGlyThrGlyAlaIleAspProGlyLys 475
1042 CACTATTCCTACCTTATAGATTAAACACGACCGCTCAATGTGTT.. 1089
    |||:::|||||:::
476 AsnSerGluLeuLysThrThrIleuThrAsnThrThrIleSerAsnTyLe 492
1090 .....AATGTTCTTTATCCGAGACAGACAGACAGAAC 1120
492 uLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr... 507
1121 CTGTTTATCATGCTGCAGGTGGGTCACAGTTATCCACCCAGCTGAAT 1170
    |||:::|||||:::
508 .....ValAsnSer.....SerIleAsn 513
1171 AATGAGAAAATATTCTTATATGCAAAAGAAAGTGAATTGATNCT 1220
    |||:::|||||:::
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1221 TACGAGCAATCAACCAAGCGCGGCGGTTGTTATTTAGGGTAAAT 1270
    |||:::|||||:::
527 .....GlnArgGlyGlyValGlnIleAspGlyAspI 538
1271 TTACGGCTCGCCCTAAAAACAGCAAAACGAGCGCGCGGCTTCAT 1320
    |||:::|||||:::
538 IeThrSerLysGlyGlyAsnLeuThrIleThrSerGlyGlyTrpValAsp 554
1321 ATCAGTGATGGCAGTACCGTTACTTGAAGTAACGGCGTGGCAACGA 1370
    |||:::|||||:::
555 ValHisLysAsnIleThr..... 560
1371 CCGCCTGTCAAAATCGCAAAAGGACGCTGCTGTTCAAGCC..... 1413
    |||:::|||||:::
561 .....LeuAspGlnGlyPheLeuAsnIleThrAlaIleAsnSerV 573
1414 .....AAAGGGAAAACCAAGCGCTGCTGAGCGTGGCGAGCGGTAAA 1455
    |||:::|||||:::
573 AlaIlePheGlnGlyGlyAsnAsnLysAlaArgAspAlaIleAsnAlaLys 589
1456 GTCATCTTAGTCAGCAGGGGAGCATCAAGCAAAAACAGCCTTTAG 1505
    |||:::|||||:::
590 IleVal..... 591
1506 TGAATCGCGCTGGTGCACGCGAGGGGAGCGTGCACATGATGCCGATA 1555
    |||:::|||||:::
592 .....AlaGlnGlyThrValThrIleThrGlyGluG 602
1556 ATCAGTTCAACCCGACAACTATTTTCGGCTTTCGCGGAGCGAGCTTGG 1605
    |||:::|||||:::
602 LysLys.....AspPheArgAlaAsnAsnVal 610
1606 GATTTGAACGGGCTTGGCTTGGTTCACCCGATTCAAAATACCGATGA 1655
    |||:::|||||:::
611 SerLeuAsnGly.....ThrGly 617
1656 AGGGGCGATGATGTC.....AACCAATCAAGACAAAGAAAT 1693
    |||:::|||||:::
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACCGTTACCATTTACAGCAATTAAGATTTACTACACCGGCAATTAAC 1743
    |||:::|||||:::
634 LysThrIleAsnIleSerGlyAsnIleThrIleAsnIleThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerTyTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667
1747 .....AACTGGAT..... 1755
667 AlaLeuAsnLeuGluThrGlyAlaAsnPheThrPheIleLysTyIleSerS 684
1756 .....ACGAAAAAGAAATGGCC 1773
684 eAsnSerLysGlyLeuThrThrGlnTyArgSerSerAlaGlyValAsn 700
1774 TACACGCTTGTTGGTGGCGAAGATGCAACCAACAGACAGCGCGCT 1823
    |||:::|||||:::
701 PheAsnGly.....ValAsnGlyAsnMe 708
1824 CAATCTGAATTTACACCGCAAGAACGGGATCGACTTACTGCTTCCG 1873
    |||:::|||||:::
708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724
1874 GCGGAACAAATTTAAC.....GGC 1893
724 roAsnGluAsnMeIleAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATCAGCGCAACA...AACGGCAACGTTTTCAGCGGCGACCGCAC 1940
    |||:::|||||:::
741 AsnIleThrAlaThrGlyGlySerValAlpheAsp..... 753
1941 ACCGCAACGCTTAC...AATCATTTAGAACGGGCGGTCAAAAATGGAAG 1987
    |||:::|||||:::
754 .....IleTyAlaAsnHisSerGlyArgGly.....AlaG 764
1988 GTATCCCAACAGAGAAATCGTGGGACACAGCATTTGATCGACCGACA 2037
    |||:::|||||:::
764 IuLeuLysMetSerGluIleAsnIleSerAsn..... 774
2038 TTTAAAGCGGAAACTTC.....CATATTCAGGCGGACCAAGC 2075
    |||:::|||||:::
775 .....GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGTTTCCCGCAATGTTGCCAAAGTGGAGAGCGGATTTGCAATTAAGA 2125
    |||:::|||||:::
789 aPhe.....LysIleAsnLysAspLeuThrIleAsn. 799
2126 ATCAGCGCCAGACAGATTTCGGTGTGCGACCGCATCAAAAGCCACACAATC 2175
    |||:::|||||:::
799 ..... 799
2176 TGTACACGCTTGCGACTGACGCGGTCTGCACAACTTACCGAAAAAACAT 2225
    |||:::|||||:::
800 AlaThrAsnSerAsnPhe..... 805
2226 TACCGACGATTAAGTGATGTTTCATTTAGCAAGACCGCATCAGAGGA 2275
    |||:::|||||:::
806 .....SerLeuArgGlnThrLysAspAspPheT 815
2276 AATGTCAGCTTCCCGATCCAGCTCATTTAAATCTCACAGAGATGCCACA 2325
    |||:::|||||:::
815 yfAspGlyTyfAlaArgAsnAlaIleAsnSerThrTyfAsnIleSerIle 831
2326 CTCAAAGGCAATCTTAGTCAGCGGAGACAGCACTATACGTTACGCG 2375
    |||:::|||||:::
832 LeuGlyGlyAsnValThrLeuGlyGly..... 840
2376 CAACGCAACCCAAACGCAACCTCAGCCTGTCGGCAATGCGCCAGCAA 2425
    |||:::|||||:::
841 .....GlnAsnSerSerSerIleThrLysn.....IleT 852
2426 CATTTAATCAAGCC.....ACATTAACGGGCAACACATCGGCTTCG 2466
    |||:::|||||:::

```

852 hrilegluylaslaalaasnvalthrleuglualaasnalaasna 868
2467 GACAAAGTTCATTATTAATCTAGACAAACGCGCTACAAAAGCGCACTCT 2516
869 GlnGln.....AsnIleArgAspArgValIleIleIleuGlySerLe 882
2517 GACGCTTTCGACAAAGCTAGGCAACGTAAGCCATTCCGCACTCAACG 2566
882 uleuValasnIleuSerLeuSerLeuThrIleGluAsnAlaSplLeuYsg 899
2567 GCATGTCTCCCTAGCCGATAGGCAGTATTCAT.....TTTGAA 2607
899 LysnleuThrIleSerIleuSerIleThrIleThrIleHisAla 915
2608 AACGCGCCCTTACCGGAAAAATACGGCGGCAAGATACGGCATACCA 2657
916 ThrLeuasnIleThrIleGlyAsnIleThrAsnAsnIleThrAlaGluIleAs 932
2658 CTTAAAAGACAGCAATGACGCTGCCGTGCGGCAAGCAATTAAGCAATT 2707
932 nileThrIleGlnIleValIleValIleuGlyAsnValIleThrAsnAspIleYsPl 949
2708 TAAACCTTGACAAAGCCACCTTACACTCAATCCGCTATCGACAGAT 2757
949 euAsn.....IleThrIleHisAla..... 955
2758 GCGCGAGCGCGCAACCGCAGTGGCGAGATGCGCGCGCGCTTC 2807
956LysArgAs 958
2808 GCGCGCTTCCTTATTAATCCGTTACGCGCAACTTCGGCAAGATCCGCTT 2857
958 nglArgSerIle..... 962
2858 TCAACAGCTGACGCTAAAGGCAATTGACGCTCAGGAGACATTCCGC 2907
963IleIleGlyIleAsnIleIleAsn..... 969
2908 TTTATGTCGAACCTTCGCTACCGCAGCGCAAAATGAGTGGCGGA 2957
970LysIleGlySerLeuAsnIleThrAs 978
2958 AAGTTC.....G 2965
978 pSerAsnAsnAspAlaGluIleGlnIleGlyIleAsnIleSerIleuYsg 995
2966 AAGGACTTACACTTGGCT.....GTCACAATTAACGGCAACGAA 3006
995 IuGlyAsnIleuThrIleSerIleAspIleAsnIleThr..... 1008
3007 CCCGTAAGCTCGAGCAATGACGTA.....GTGAAGGAAAAA 3047
1009LysGlnIleThrIleIleYsIleGlyIleAspIleGluIle 1021
3048 C...AACACACCGCTGTCCGAAATCTTAATTTACCTCGAAAACGAA 3094
1021 pSerIleSerIleAspAlaThrIleSerIleAsnAlaAsnIleuThrIleYs..... 1035
3095 ACGTCGATGCGCGCATGGCTATATCGCAAGACGGCGAG 3144
1035 1035
3145 TTCCGCTGCAATAATCCGGTCAAGAACAGAGTTCCGCAACATCGG 3194
1036ThrIleGluIleuThrIleuThrIleuAspLeuSe 1046
3195 CAAGCGGGAAGAACAGAGCGCTTGACGGCAAAACAGGCAACATCTG 3244
1046 rIleSerIleYsAsnIleYsAlaGluIleThrAlaIleYsAspIleYsAspIle 1063
3245 CCGCAACACAGCGGCGGAAAAAGCAACGCGCAAGC..... 3282
1063 euThrIleGlyIleAsnIleSerIleAsnAspIleYsAsnIleGlyIleAlaIleYsThr 1079
3283CTGACGCGCTGATTGGCGC...GGCGCA 3311
1080 ValThrPheAsnAsnValIleYsAspSerIleIleSerAlaAspIleHisAs 1096
3312 TGGCACGGAAGAAAGCAGAAAGTGTGCGAACCGCGCGCGGCAAGCGG 3361
1096 nValThr.....LeuAsnIleYsValIleYsThrIleSerIleSerIle 1111
3362 GGGAAATATCCCGCATTAATGACGGCGAGAGAGAAAAAACGGGTGCG 3411
1111 LysArgIleuSer.....Asn 1115
3412 GCGGATTAAGACACGCGCTTG.....GGAAACAGCGGAGCGGA 3452
1116 SerAspAsnAspIleThrIleuThrIleThrIleAlaIleYsAsnValGluVal 1131
3453 AACCGCGCGCTACCCAGCGCTTCCCGCGCGCGCGCGCGCGG 3502
1131 1131
3503 ATTGCGGCAACCGGACGCCCAACCGCAACCCACCGGCGGCGAGCTG 3552
1132AsnIleYsAspIle 1135
3553 ATCAGC...CGTTATGCAATAGCGGTTTGAATTTTCGCGCAGCT 3599
1136 ThrSerLeuIleThrValAsnIleThrIleAlaSerIleGluYsValIleThr.T 1152
3600 CAACAGCGTTTCGCGCTACAGAGCAATGACCGCGCTTGGCGGAAG 3649
1152 hrAlaGlySerThrIleAsnAlaThrAsnIleYsAlaSerIleThr 1168
3650 ACCGCGCAACGCGCTTGACAAAGCGGATCCGGACACCAACATCAC 3699
1169 LysThrGlyAspIleSerIleThrIleSerIleGlyAsnThrValSerValSe 1185
3700 CGTTGCG...AAGATTCCGCGCTACCGCAACAAACCGACCTGGCCA 3746
1185 rAlaThrValAspLeuThrIleYsSerIleYsIleGluAlaIleYs 1202
3747 AATCGGTATGCAAAAAACCTCGCGCGCGCGCGCTGCGCATCGTTT 3796
1202 erIleGluIleAsnValIleThrSerAlaThrIleIleYsIleYs 1216
3797 CGCACAACGCGACCGGAACACCTTCGACGAGCATCGCA..... 3838
1217 ...ThrIleSerIleYsAsnThrValAsnValIleThrAlaAsnAlaGlyAspLe 1232
3839ACTGCGCACGCGCTTGGCCACGG 3860
1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrIleGluIleAlaIleThrL 1249
3861 TGCCGTTTCGGGCAAT.....ACGGCATCGGAGGTTTCGACATCGCA 3904
1249 euThrAlaThrIleGlyAsnThrIleuThrIleGluAlaIleYsSerIleThr 1265
3905 TCAGCGCGGCGCGGCTTGAAGCGGCAAGCGCTTTCAGACGCAATCAGA 3954
1266 SerThrIleYsGlyIleValIleAspLeuAla.....GlnAsnIleYsIle 1280
3955 GGCAAAATCCGCGCGCGCTGCTGCTTACGGCATTCAGGCAAGATACG 4004
1280 eAlaGlySerIleAsnAlaIleAsnValIleThrLeuAsnThrIleGlyThrL 1297
4005 CCGAGGTTTCGGGATTCGGCATCGAACCGCAACCGGCGCAAGCGCT 4054
1297 euThrThrValAlaGlySerAspIleYsAlaThrSerIleYsThrIleuVal 1313
4055 ATTTCGTCCAAA 4066
1314 IleAsnAlaIleYs 1317


```

COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-APR-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-617-697-4

```

```

alignment_scores:
Quality: 274.50 Length: 1488
Ratio: 0.419 Gaps: 73
Percent Similarity: 44.019 Percent Identity: 19.220

```

```

alignment block:
US-09-303-518D-653 x US-08-617-697-4

```

```

Align seq 1/1 to: US-08-617-697-4 from: 1 to: 1477

```

```

178 AATAAGGCAAGTTGTCAGTCGGCGGCAAGATATTGAGTTTAAACAA 227
||| ||| ::::::::::||| ||| ||| ::|||
199 AsnHisGlyLeuIleThrValGly..LysAspLysVal..... 211
228 AAAAGGAGTGTGTCGCAATCGATGACGAAACCCCGATGATTGATT 277
:::|||||::: ||| ::::: |||
212 .....AsnLeuIleGlyLysValLysAsnGlu..... 221
278 TTTCGTGTCGTCGTAACGGC.....GCGGCGCATTTGGCGCGCAT 321
:::|||||::: ||| ||| ||| ||| ||| ||| |||
222 ..GlyValIleSerValAsnGlyLysSerIleSerLeuLeuAlaGlyGln 237
322 CAA.....TATATTGTGAG 335
::: |||
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIleAla 254
336 CGTGCAATACGCGGCTATACCAATGTTGTTGGTGGGAGGAA 365
::: ||| ||| ||| ||| ||| ||| ||| ||| |||
254 AlaProGluAsnGluAlaValAsnLeuGlyLysPheAlaLysGly 271
386 GCAAT.....CCGATCAGCACCGCTTTTCT 411
::: ||| ||| ||| ||| ||| ||| ||| ||| |||
271 LysAsnIleAsnValArgAlaIleThrIleArgAsnGlnGlyLysLeuSer 287
412 TACCAAAATGTGAAGAAATATATATTAAGACGAGACTAAGCGCATCC 461
::: ||| ||| ||| ||| ||| ||| ||| ||| |||
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTATGCGGCGCATATATATATGCGGCTTTGCAACAATTTGTCAAGATG 511
::: ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

298 .....AsnIleValLeuSer 303
512 CAGAACCTGTTGAG.....ATGACCATGTTATATGAT 543
||| ||| ::::::::::|||
303 LysGluGlyGluAlaIleGlyValIleSerAlaGlnAsnGln 319
544 GGGTGAATAACGCTGATTTAAATAATACCGTATGTT.....CG 587
||| ||| ||| ||| ||| ||| ||| ||| |||
320 GlnAlaLysGlyLysLeuMetIleThrGlyAspLysValThrLeuLys 336
588 AATCGAGCA.....GCGACACAAATAT 610
::: ||| ||| ||| ||| ||| ||| ||| ||| |||
336 ThrGlyAlaValAlaIleAspLeuSerGlyLysGluGlyLysIleThrLys 353
611 GCGGCTGATGAA...CACGAACCAATATAC..... 639
::: ||| ||| ||| ||| ||| ||| ||| ||| |||
353 euGlyLysAspLysLysGlyLysAsnGlyLysIleGlnLeuAlaLys 369
640 .....CGCAAAATTCATATGATTTGCAAC..... 666
::: ||| ||| ||| ||| ||| ||| ||| ||| |||
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGluLys 386
667 .....GCATATCTTGG.....CTGCTGCTGCAATA 694
||| ||| ||| ||| ||| ||| ||| ||| |||
386 sGlyLysPheAlaIleValIlePheLysPheAlaLeuIleAspLysAsnI 403
695 CCTTGCACAAAATGATCAGTGTGGTGCACAGTCACCTTAGTAGAGCAA 744
||||| ||| ||| ||| ||| ||| ||| ||| |||
403 LeuAsnAlaGln...GlySerGly..... 409
745 AAAATTAACATAGCCCATATGCTTTTACCACAGAGCTCATTTGG 794
||| ||| ||| ||| ||| ||| ||| ||| |||
410 .....AspIleAlaLysThrGlyLysPheValGln 419
795 CGACAGTGGCTCACCATGTTATCTATGATGCCCAAAAGCAAGTGT 844
||||| ||| ||| ||| ||| ||| ||| ||| |||
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGTATTCACAAACAGCAACCCCTATATAGAAAAGCAAT 894
||| ||| ||| ||| ||| ||| ||| ||| |||
430 .....AsnAlaIleValAspAla..... 435
895 GCGTTCACAGTATGTCGTAAGATTGTTCTATGATGAAATCTTTGCTGG 944
||| ||| ||| ||| ||| ||| ||| ||| |||
436 .....LysGluThrLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGTATCTACGAA...CCACATCAAAATGGAATACT 991
||| ||| ||| ||| ||| ||| ||| ||| |||
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGlyI 459
992 TTTTAAACGACATATATATGCGCGAGGAAAATGATGCCAAACATATA 1041
::: ||| ||| ||| ||| ||| ||| ||| ||| |||
459 LeuAsnAspLysPheProThrGlyThrGlyLysGluAlaSerAspProLys 475
1042 CACTATTCCTACCTTATATGATTAATAACACGACACCGTTCATTTT 1089
::: ||| ||| ||| ||| ||| ||| ||| ||| |||
476 AsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnThrLe 492
1090 .....AATGTTCTTATTCGACGACAGCAAGCAAGAAC 1120
||| ||| ||| ||| ||| ||| ||| ||| |||
492 uLysAsnAlaIlePheIleAsnIleThrAlaSerArgLysLeuThr.... 507
5121 CTGTTATCATGCTCAGGTGGGTCACAGTATATGACCCAGACTGAT 1170
||| ||| ||| ||| ||| ||| ||| ||| |||
508 .....ValAsnSer.....SerIleAsn 513
1171 AATGAGAAATATTTCTTATTTGACAAAGAAAGGATTTATATCT 1220
||| ||| ||| ||| ||| ||| ||| ||| |||
514 IleGlySerAsnSerIleLeuIleuHisSerLysGly..... 526
1221 TACACACACATCAACACAGCGCGGCTTTGATTTTGAAGGTAAAT 1270
||| ||| ||| ||| ||| ||| ||| ||| |||
527 .....GlnArgGlyLysGlyValGlnIleAspGlyAspI 538

```

```

1271 TTACGGCTCGCTAAACAAACGAAACGTGGCAGCGCGCGCTCAT 1320
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
538 lenthSerlysglyAlaSerLeuThrIleTyrSerGlyIleTyrPalaasp 554
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1321 ATCAGTGAAGGAGTACCGCTTACTTGAAGTAAGCGCGCTGGCAACGA 1370
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
555 ValHisIleAsnIleThr..... 560
1371 CCGCGCTGCCAAATCGCAAGGACGCTGCTGTTCAAGC..... 1413
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
561 .....LeuAspGlnGlyPheLeuAsnIleThrAlaAlaSerV 573
1414 .....AAGGGGAAACCAAGGCTCGCTGCGTGGCGAGCGTAA 1455
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
573 AlaIaPheIleGlyAlaSerAsnIleValArgAspAlaIaSerAlaIys 589
1456 GTTCATCTTGAATCAGCGGAGCATCAAGCAAAAAAAGCCTTTAG 1505
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
590 IleVal..... 591
1506 TGAATATCGCTTGTTCAGCGGAGCGGAGCATCAATGATGCGGATA 1555
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
592 .....AlaGlnGlyThrValThrIleThrGlyGlu 602
1556 ATCAGTTCAACCCGACAAACTTATTTGCGCTTCGCGGAGCGCTTTG 1605
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
602 Lys.....AspPheArgAlaSerAsnVal 610
1606 GATTTACAGCGGATTCGCTTCCTTCACCGCATTCAAATACGATGA 1655
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
611 SerLeuAsnGly.....ThrGly 617
1656 AGGGGAGATGATGTC.....AACCATCAAGCAAGAT 1693
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
617 SerGlyLeuAsnIleIleSerSerValAsnSerLeuThrHisAsnLeuSerG 634
1694 CCACCGTTACCATTTACAGCATTAAGATATTACTACACCGCATTAAC 1743
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
634 LysThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC..... 1746
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
651 AsnThrSerTyrTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667
1747 .....AACTGGAT..... 1755
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
667 AlaAsnLeuGlnThrGlyAlaAsnPheThrPheIleLysTyrIleSerS 684
1756 .....AGCAAAAAAGAAATGGC 1773
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
684 eraSerSerLysGlyLeuThrThrGlnTyrArgSerSerAlaGlyVala 700
1774 TACACGGTTGGTTGGCAGAAAGATGCACCAAAACGACGGCGGCT 1823
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
701 PheAsnGly.....ValAsnGlyAsnMe 708
1824 CAATCTGAATACCAACCGGAGAGGATGCGACTTTCCTTCCTTCG 1873
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
708 tSerPheAsnLeuLys...GlnGlyAlaLysValaSerPheLysLeuLysP 724
1874 GCGGAACAAATTTAAAC.....GGC 1893
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
724 roAsnGlnAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATCAGCAACA...AACGCAAACTTTTTCAGCGGAGCGGAC 1940
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
741 AsnIleThrAlaThrGlyGlySerValPhePheAsp..... 753
1941 ACCGACGCGCTAC...AATCATTTAGAGCGGCTGTCAAAAATGAAG 1987
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
754 .....IleTyrAlaAsnHisSerGlyArgGly.....AlaG 764

```

```

1988 GTATCCCAAGAGAAATCGTGTGGACACAGATTGATCGACCGACA 2037
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
764 IuleuLysMetSerGluIleAsnIleSerAsn..... 774
2038 TTTAAAGCGGAAACTTC.....CATATTACGGCGGACAGC 2075
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
775 .....GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGGTTCCCGCAATGTTCGCAAGGTGAAGCGGATTTGGCATTTAGCA 2125
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
789 aPhe.....LysIleAsnLysAspLeuThrIleAsn. 799
2126 ATCAGCGCCACAGAGTTTTCGTGTCCACCGCATCAAGCCACACAATC 2175
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
799 ..... 799
2176 TGTACAGCTTGGAGCTGAGCGGCTGTACAAAGTTTACCCAAAAACAT 2225
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
800 AlaThrAsnSerAsnPhe..... 805
2226 TACGACGATTAAGTATGCTTCTTCATTCAGCAAGCCGACATCAGAGCA 2275
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
806 .....SerLeuArgGlnThrLysAspAspPheT 815
2276 ATGTCAGCTTCGCCGATCAGCTCATTTAAATCAGAGACTTGCACA 2325
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
815 YrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 831
2326 CTCACAGGCATCTTAGTCAGCGGAGACACGACATACGTTACGGC 2375
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
832 LeuGlyGlyAsnValThrLeuGlyGly..... 840
2376 CAACGCCACCCAAAGCGCAACTGAGCTGTGGCAATGCCACGAA 2425
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
841 .....GlnAsnSerSerSerSerIleThrGlyAsn.....IleT 852
2426 CATTTAATCAAGC.....ACATTAAACGCGACACATCGGCTTCG 2466
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
852 ThrIleGlyLysAlaAlaAsnValThrLeuGlnAlaAsnAsnAlaProAsn 868
2467 GACATGCTTCAATTAATCTAAGCAACGCGGATCAAAACGCGAGCT 2516
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
869 GlnGln.....AsnIleArgAspArgValIleLysLeuGlySerIle 882
2517 GACGCTTCGACAGACGCTAAGCAAGTAAGCATTCGCGACTCAAGC 2566
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
882 IuleuValAsnGlySerLeuSerLeuThrGlyGlnAsnAlaSerIleLysG 899
2567 GCAATGCTCCCTAGCCGATTAAGCAGTATTCAT.....TTGAA 2607
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
899 LysAsnLeuThrIleSerGlnSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCCGCTTACCGGAAATCAGCGGCGCAAGATAGCGCATTTCA 2657
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
916 ThrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGlnIleAs 932
2658 CTTAAAGACAGCAATGAGCGCTCCGTCGCGCAAGCAATTAAGCAAT 2707
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
932 rIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspL 949
2708 TAAACCTTGACAGCCACCATTAATACATCAATTCGCGCTATCGACAGAT 2757
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
949 euAsn.....IleThrThrHisAla..... 955
2758 GCGGACGCGCGCAACCGGACGTCGGGAGATGCGCGCGCGCGCTTC 2807
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
956 .....LysArgAs 958
2808 GCGCGCTCCCTATTATCGTTACCGCGCAACTTCGCGAGAAATCCGTT 2857
    :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
958 nGlnArgSerIle..... 962
2858 TCAACACGCTGACGTTAAAGCAAAATTAAGCGGTACAGGAAACATTCGC 2907

```


FILING DATE: 16-MAR-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Beirstresser, Jerry W
 REGISTRATION NUMBER: 22,651
 REFERENCE/DOCKET NUMBER: 1038-625
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1477 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-719-641-4

alignment_scores:
 Quality: 274.50 Length: 1488
 Ratio: 0.419 Gaps: 73
 Percent Similarity: 44.019 Percent Identity: 19.220

alignment block:

US-09-303-518d-653 x US-08-719-641-4 ..

Align seg 1/1 to: US-08-719-641-4 from: 1 to: 1477

```

178 AATAAGGAGAGTTGCTGCTGCGGCAAGAAATATTGAGTTTACACAA 227
    ||| ||| : : : : : ||| ||| ||| ||| : : : : :
199 AsnHisGlyLeuIleThrValGly..LysAspGlySerVal..... 211
228 AAAAGGAGAGTTGCTGCGCAATCGATGACGAAGCCCGATGATGATT 277
    : : : : : ||| : : : : : : : : : : : : : : :
212 .....AsnLeuIleGlyGlyValValLysAsnGlu..... 221
278 TTTCTGTGATCGCTTACGCGC.....GTGCGGCGATGCGGCGCAT 321
    : : : : : ||| ||| ||| ||| : : : : : ||| ||| |||
222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGly 237
322 CAA.....: : : : : : : : : : : : : : : : : : : :
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIle 254
336 CGTGCGACATAACGCGCGCTATACAAATGTGATTTGGTGGCGAGGAA 385
    : : : : : ||| : : : ||| : : : : : ||| : : : : :
254 AlaIleProGluAsnGluAlaValAlaSerLeuGlyAspIleAlaLys 271
386 GCAAT.....: : : : : : : : : : : : : : : : : : : :
271 LysAsnIleAsnValArgAlaIleThrIleArgAsnGlnGlyLysLeu 287
412 TACCAATGTGAAAGAAATATATATAAGACGAGCTAACGCGCATCC 461
    : : : : : ||| : : : ||| : : : : : ||| : : : : :
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTATGCGGCGATTCATATATCCGCGTTTGCACAAATTTGTACAGAT 511
298 .....: : : : : : : : : : : : : : : : : : : :
512 CAGAACCTGTGAG.....ATGACGAGTTATATGAT 543
    : : : : : ||| : : : : : : : : : : : : : : : :
303 LAlysGlnGlyGluAlaGlnIleGlyGlyValIleSerAlaGlnAsnGln 319
544 GGGTGAATACGCTGATTAATAATACCTGATCGTGT.....CG 587
    : : : : : ||| : : : ||| : : : : : ||| : : : : :
320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeu 336
588 AATCGAGGAGCA.....GGCAGACATATAT 610
    : : : : : ||| ||| : : : : : ||| : : : : :
336 sThrGlyAlaValIleAspLeuSerGlyLysGlnGlyGlnThrTyrL 353
611 GCGGCTGATGAA..GACGAACCCCAATATAC..... 639
    : : : ||| ||| : : : |||
  
```

```

353 euGIyGlyAspGluArgGlyGlnGlyLysAsnGlyIleGlnLeuAlaLys 369
640 .....CGGGAAGTTCAATCATATTCGCAAGC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGly 386
667 .....GCATATTCCTTG.....CTGCGTGGGCAATA 694
    ||| ||| : : : : : ||| : : : : :
386 sGlyGlyPheAlaIleValTrpGlyAspIleAlaLeuIleAspLysAsn 403
695 CCTTGGCAAAATGATCATAGTGTGGTGGCACGTCAACTTAGTAGCGAA 744
    ||| ||| : : : : : ||| : : : : :
403 LeAsnAlaGln...GlySerGly..... 409
745 AAATTAACATAGCCCATATGTTTTCACACAGAGGCGTATTTGG 794
410 .....AspIleAlaLysThrGlyGlyPheValG 419
795 CGACAGTGGCTACCAATGTTATCTATGATGCCCAAGCAAAAGTGT 844
    ||| ||| : : : ||| |||
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGCTATTTGCAACACAGCAACCCCTATATGCAAAAGCAAT 894
    ||| : : : : : : : : : : : : : : : :
430 .....AsnAlaIleValAspAla..... 435
895 GCGTTCCAGCTAGTTGTAAGATTGGTTCATGATGAAATCTTGTCTG 944
436 .....LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGTATTTCTACGAA...CCACATCAAAATGGAAATACT 991
    ||| : : : : : ||| : : : : : ||| : : : : :
442 eAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThrGly 459
992 TTTTAAACGACAAATATATATGCGCGAGGAAATATCATGCCAACAATAA 1041
    : : : : : : : : : : : : : : : : : : : :
459 LeAsnAspGluPheProThrGlyThrGlyAlaIleAsnAspProLys 475
1042 CACTATCTCTACCTTATAGATTAAACAGACGCGTCAATGTGT... 1089
476 AsnSerGluLeuLysThrThrLeuThrAsnThrIleSerAsnTyrLe 492
1090 .....AATGTTCTTTATCCGACAGACAGCAAGAGAAC 1120
492 uLysAsnAlaIleTrpHisMetAsnIleThrAlaSerArgLysLeuThr... 507
1121 CTGTTTATCATGTCGACAGTGGGTCAACAGTTATGACCCAGCATGAAT 1170
508 .....ValAsnSer.....SerIleAsn 513
1171 AATGAGAAATATATTCCTTATTTGACAAAGAAAGTCAATGATTA 1220
    ||| : : : : : ||| : : : : : ||| : : : : :
514 IleGlySerAsnSerHisLeuIleLeuHisSerLysGly..... 526
1221 TACACGAAACATCAACCAAGCGCGGCGGTTTATTTGAGGATAT 1270
527 .....GlnArgGlyGlyGlyValGlnIleAspLysAspI 538
    : : : : : ||| : : : ||| : : : : : ||| : : : : :
538 IeThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAsp 554
1321 ATCAGTATGCGCATACCGTTACTTGAAAGTAAACGCGGTGCAACGA 1370
    : : : : : ||| : : : : : ||| : : : : :
555 ValHisLysAsnIleThr..... 560
1371 CCGCCTGTCGCAAAATCGGCAAGGACGCTGCTGTTCAAGCC..... 1413
561 .....LeuAspGlnGlyPheLeuAsnIleThrAlaIleAsnSer 573
1414 .....AAAGGGAAACCAAGGCTGCTGACGCTGGCGGACGATAA 1455
    : : : ||| ||| : : : : : ||| : : : : :
573 AlaIlePheGlnGlyLysAsnAsnLysAlaIleArgAspAlaIleAsnLys 589
  
```

1456 GTCATCTTAGATCAGACGGCGAGATCAAGCAAAAAACAAGCTTTAG 1505
1590 IleVal..... 591
1506 TGAATCGCGTGGTCAGCGGAGGAGGAGCGTGCAACTGATGCCGATA 1555
592AlaGlnGlyThrValThrIleThrGlyGluG 602
1556 ATCAGTCAACCCGCAAACTCATTTGGCTTGGCGGCGGAGCGCTTG 1605
602 Lys.....AspPheArgAlaAsnVal 610
1606 GATTGACGAGGCAATTCGTTGCTCCACCGCATTCAAATACCGGATA 1655
611 SerLeuAsnGly.....ThrGly 617
1656 AGGGCGCATGATGTC.....AACCAATCAAGACAAGAAAT 1693
617 sGlyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerG 634
1694 CCACCGTTACCATTCAGCAATTAAGATATTACTACACCGGCAATTAAC 1743
634 LysThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThrThrArgLys 650
1744 AAC..... 1746
651 AsnThrSerLysTrpGlnThrSerHisAspSerHisTrpAsnValSerAl 667
1747ACTTGAT..... 1755
667 AlaLeuAsnLeuGlnThrGlyAlaAsnPheThrPheIleLysTrpIleSerS 684
1756AGCAAAAAAGAAATGGC 1773
684 eAsnSerLysGlyLeuThrThrGlnTrpArgSerSerAlaGlyValAsn 700
1774 TACACGCTGGTGGTTGGCGAAGAAATGCAACCAAGCAAGCGGCGCT 1823
701 PheAsnGly.....ValAsnGlyAsnMe 708
1824 CAATCTGAATTTACCAACCGGAGAGGAGTGCACCTTACTGCTTCCG 1873
708 tSerPheAsnLeuLys...GluGlyAlaLysValAsnPheLysLeuLysP 724
1874 GCGGAAACAATTTAAAC.....GGC 1893
724 roAsnGlnAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla 740
1894 AATATCAGCAACA...AACGGCAACTGTTTTCAGCGGAGACCGGAC 1940
741 AsnIleThrAlaThrGlyGlySerValPhePheAsp..... 753
1941 ACCGACGCGCTAC...AATCATTTAGAGACGGGTGTCAAAAATGGAAG 1987
754IleTyrAlaAsnHisSerGlyArgGly.....AlaG 764
1988 GTATCCCAAGAGGAGAAATCGTGGGAGCAACGATTGGATGACCGGACA 2037
764 LuLeuLysMetSerGluIleAsnIleSerAsn..... 774
2038 TTTAAACGGGAAACTTC.....CATATTCAGGCGGAGACAAGC 2075
775GlyAlaAsnPheThrLeuAsnSerHisValArgGlyAspAspAl 789
2076 GGTGGTTTCCCGCAATGTGCCAAAGTGAAGCGGATGGCATTTAAGCA 2125
789 apHe.....LysIleAsnLysAspLeuThrIleAsn. 799
2126 ATCAGCGCCAAGAGCTTTCGGTGTCCACCGGATCAAAACCAACAATC 2175
799 799

2176 TGATACAGTTTCGAGCTGGACGGGTCTGACAAGTTGTACGAAAAAACCAT 2225
800 AlaThrAsnSerAsnPhe..... 805
2226 TACCCAGCATTAAGTATTGCTTCTCATTTGACGAAGACGACATCAGAGCA 2275
806SerLeuArgGlnThrLysAspAspPheT 815
2276 ATGTCAGCCTTCGCGATCAGCGCTCATTTAAATCTCAGACGACTTGCCA 2325
815 YrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIle 831
2326 CTCACGGCAATCTTAGTCAGCGGAGACACGCACTATACGTTACGCG 2375
832 LeuGlyGlyAsnValThrLeuGlyLys..... 840
2376 CAACGCCACCAAAAGCGCAACTCAGCTGTGGGCAATGCCAAGCA 2425
841GlnAsnSerSerSerIleThrIleLysAsn.....IleT 852
2426 CATTTAATCAAGCC.....ACATTAACGGCAACACATCGGCTTCG 2466
852 hrIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAlaProAsn 868
2467 GACAAATGCTTCAATTAATCTAAGCAACACGCGGTACAAAACGCGACT 2516
869 GlnGln.....AsnIleArgAspArgValIleLysLeuGlySerLe 882
2517 GACGCTTTCGCAACGCTAAGGCAACGTAAAGCATTCGCCACCTCAACG 2566
882 uLeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysG 899
2567 GCAATGTCCTCCCTACCGCATAGCGATATCCAT.....TTTGA 2607
899 LysAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAsp 915
2608 AACAGCGCTTTACCGGAAAAATACGCGCGGCAAGATACGGCATTTCA 2657
916 ThrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAs 932
2658 CTTAAAGACGCAATGAGCGCTGCGGCGGCGACAGATTAAGCAAT 2707
932 nIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspL 949
2708 TAACTTGCACAACGCCACCATTCATCAATTCGCCCTATGCACACGAT 2757
949 euAsn.....IleThrThrHisAla..... 955
2758 GCGGACGCGCGCAACCGGACGTGCGGACATGCGCGCGCGCGCTTC 2807
956LysArgAs 958
2808 GCGCGTTCCTCATTTATTCGCTTACCGCGCAACTTCGCGAGATCCGTT 2857
958 nGlnArgSerIle..... 962
2858 TCACACGCTGACGTTAAACGGCAATTTGAACGCTCAGAGAAATTCGCG 2907
963IleGlyGlyAspIleIleAsn..... 969
2908 TTTATGTGGAACTCTTGGCTACCGGACGCGCAATTTGAAGCTGGCGGA 2957
970LysLysGlySerLeuAsnIleThrAs 978
2958 AAGTCC.....G 2965
978 pSerAsnAsnAspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysG 995
2966 AAGCACTTACACTTGCT.....GTCAACAATACCGGCAACGAA 3006
995 LuGlyAsnLeuThrIleSerSerAspLysIleAsnIleThr..... 1008
3007 CCCGTAAGTCTGAGCAATTCAGCGTA.....GTGAAGGAAAAAGA 3047

```

1009 .....LysGlnIleThrIleLeuLysGlyLeuAspGlyIleAs 1021
3048 C...AACACACCGCTGTCGGAATATTTACCCCTGCAACAGAAC 3094
1021 pSerSerAspAlaThrSerSerAlaAsnLeuThrIleLys..... 1035
3095 ACCTGATGCGCGGCGATGCGTTATCAGCTTATCCGCAAGACGCGAG 3144
1035 ..... 1035
3145 TTCGCGCTCATATTCGCTCAAGAACAGACCTTCCGACAACTCGG 3194
1036 .....ThrLysGluLeuLysLeuThrGluAspLeu 1046
3195 CAAGCGCGGAGAAACAGAGCGCGCTTGGACGCAAAACAGCACACTG 3244
1046 rIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAsp 1063
3245 CCGCGCAACACAGCGCGGAAAGACACGCGCAAGC..... 3282
1063 eutThrIleGlyAsnSerAsnAspGlyAsnSerGlyAlaGluAlaLysThr 1079
3283 .....CTTGACGCGCTGATTGCGGCC...GGCGCGCAA 3311
1080 ValThrPheAsnAsnValLysAspSerLysIleSerAlaAspGlyHisAs 1096
3312 TGCACCGCAAAAGCAGAAAGTGTGCGGACGCGCGCGCGACGAGCGG 3361
1096 nValThr.....LeuAsnSerLysValLysThrSerSerAsnGly 1111
3362 GGGAAATGCGCGCATTTATGACGCGAGAGAGAAAGACGGTGCAG 3411
1111 LysArgLysSer.....Asn 1115
3412 GCGGATTAAGACACCGCTTG.....GCGAAACAGCGCGAGCGGA 3452
1116 SerAspAsnAspThrGlyLeuThrIleThrAlaLysAsnValGluVal... 1131
3453 AACCGCGCGGTACACCGCGCTTCCCGCGCGCGCGCGCGCGG 3502
1131 ..... 1131
3503 ATTTCGCGCAACCGACGCCCAACCCCAACCGACGCGCGACTG 3552
1132 .....AsnLysAspIle 1135
3553 ATCAGC...CGTTATGCCAATAGCGTTTGAGATTTTCCGCGCGCT 3599
1136 ThrSerLeuLysThrValAsnIleThrAlaSerGluLysValThrThr.T 1152
3600 CAACAGCGTTTTCGCGCTACAGAGCAATTTGACCGCTGTTGCCGAG 3649
1152 hrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThr 1168
3650 ACCGCGCGCAACCGCTTGGACACGCGCATCGGAGCAACCAACTAC 3699
1169 LysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSe 1185
3700 CGTTGCC...AGATTTCGCGCGCTACGCGCAACCAACCGACTCGGCCA 3746
1185 rAlaThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLys 1202
3747 AATCGGTATGCAAGAAACCTGCGACGCGCGCGCTGCGATCTGTTT 3796
1202 erGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyLys..... 1216
3797 CGACACACCGGACCGCAACACCTTGCAGACGCGCATCGCA..... 3838
1217 ...ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLe 1232
3839 .....ACTCGCGCGCGCTTGCACGCG 3860

```

```

1232 uThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaLeuThr 1249
3861 TGCCTTTTCGGGCANT.....ACGCGATCGCGAGGTTTCGACATCGCA 3904
1249 eutThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThr 1265
3905 TCAGCGCGCGCGGGTTTATGACGCGCGCTTTCAGACGATCGCA 3954
1266 SerThrLysGlyGluValAspLeuLeuAla.....GlnAsnGlySerI 1280
3955 GCGCAAAATCCGCGCGCGCTGCTGCATTCAGGATTCAGCAAGATACCG 4004
1280 eAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrL 1297
4005 CCGAGTTTCGGCGGATTCGCGATTCGACACCGCATCGCGCGCGCT 4054
1297 eutThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuVal 1313
4055 ATTTCGTCGCAA 4066
1314 IleAsnAlaLys 1317

```

seq_name: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-460-269C-4

seq_documentation_block:

Sequence 4, Application US/08460269C

Patent No. 6197548

GENERAL INFORMATION:

APPLICANT: CLARE, JEFFREY J.

ROMANOS, MICHAEL A.

TITLE OF INVENTION: EXPRESSION OF HETEROLOGOUS PROTEIN IN YEAST

NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Millen, White, Zelano & Branigan, P.C.

STREET: 2200 Clarendon Blvd., Suite 1400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/460,269C

FILING DATE: 02-Jun-1995

ATTORNEY/AGENT INFORMATION:

NAME: Lebovitz, Richard M.

REGISTRATION NUMBER: 37,067

REFERENCE/DOCKET NUMBER: Popov-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 243-6333

TELEFAX: (703) 243-6410

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 911 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-08-460-269C-4

alignment_scores:

Quality: 274.00

Ratio: 0.584

Percent Similarity: 40.818

Percent Identity: 20.540

Length: 1149

Gaps: 50

alignment_block:

US-09-303-518D-653 x US-08-460-269C-4 ..

Align seg 1/1 to: US-08-460-269C-4 from: 1 to: 911

```

1306 GCGCGGGGGCTTCATATCAGTATGCGCAGTACCGCTTACTTGCAAAATATA 1355
      |||||
58  G1YAlaG1YAlaIArgThrAlaThrGlyThrIle.....LysValSe 72
1356 GCGCGGGGCAAGACGCGCGCTGTCAAAATCGGCAAGGCGCGCGCTG 1405
      |||||
72  rGly.....ArgGlnAlaGlnGlyValLeuLeuG 82
1406 TTCAGCGCAAGGGGAAACCAAGGCTCGGTACGCTGGCGGCGAGTAA 1455
      :|||
82  LAsnProAlaAlaGlu.....LeuArgPheGlnAsnGlySer 94
1456 GTATCTTGAATCAGCAGCGGCGGCGATCAAGCAAAAAACAGCCCTTAG 1505
      |||
95  ValThrSerSerGlyLeuPheAspGluGlyValArgPheLeu... 110
1506 TGAATCGCGCTGTGATCAGCGGCGGAGCGGTCAACTGAATGCGGATA 1555
      |||||
111 .....GlyThrValThrValLysAla.... 117
1556 ATCAGTTCAACCCCGACAACTATTTGGCTTTCGCGGCGAGCTTGG 1605
      |||
118 .....GlyLysLeu 120
1606 GATTTCAGCGGCGATTCGCTTGTTCACCGCATTCMAAATACGATGA 1655
      :|||
121 ValAlaAspHisAlaThrLeuAlaAsnValSerAspThrArgAspSps 137
1656 AGGGCGCATATTTGTCAACCAATCAAGCAAGAAATCCACGTT... 1701
      |||||
137 pGlyIleAlaLeuTyrValAlaGlyGlnAlaGlnAlaSerIleAla 154
1702 ....ACCATTCAGCGCAATAAAGATATTACTCAACCGGCATTAACAC 1746
      |||||
154 spSerThrLeuGlnGlyAlaGlyValArgValGluArgGlyAlaAsn 170
1747 AACTGTGATGCAAAAAGAAATTCCTACACGCTTGGTTGGCGAGAA 1796
      :|||
171 ValThrValGlnArg..... 175
1797 AGATGCAACCAAAAGCGGCGCTCAATCTGAAT.....T 1834
      :|||
176 ....SerThrIleValAspGlyGlyLeuHisIleGlyThrLeuGlnPro 191
1835 ACCAAGCGGGAAGAGCGGATCGCACTTACTGCTTCCGCGGCAACAAT 1884
      |||||
191 eugInProGluAspLeuProPheSerArgValIleLeuGlyAspThrSer 207
1885 TTAAAGCGCAATATCAGCAAAACGCAAACTGTTTTCAGCGGCGAG 1934
      :|||
208 ValThrIle.....ValProAlaSerGly..... 215
1935 ACCGACACCGCAGCGCTTACATCATTTAGGAAGGGGTGTCAAAATGG 1984
      :|||
216 ....AlaProAlaIleValSerValPheGlyAlaAsnGluLeuThrVal 231
1985 AAGGTATCCCAAGAGAAATCGTGGGACAAACGATTGGATCGACCGC 2034
      :|||
231 spGly..... 232
2035 ACATTTAAAGCGAAACTTCATATTCAGGCGGCAAGCGGCTTTC 2084
      |||||
233 .....GlyHisIleThrGlyValArgAlaAla.... 241
2085 CCGCAATGTTGCCAAAGTGAAGCGGATGGCAATTAAGCAATCAGCGCC 2134
      :|||
242 ....GlyValAlaIleAlaMetAspGly..... 248
2135 AAGCATTTTCGGTGTGCGACCGCATCAAAAGCCACACATCTGTACGCT 2184
248 ..... 248

```

```

2185 TCGGACTGAGCGGCTGTGACAAAGTTGTACGCAAAAACCATTAACGAGCA 2234
248 ..... 248
2235 TAAAGTATGCTTCATTTAGCAAGACCGACATCAGAGCGCAATGACGC 2284
      :|||
249 ....AlaIleValHisLeuGlnArgAlaThrIleArg..... 259
2285 TTGCGGATCAGCGTCATTTAATCTCACAGGACTTGGCACACTCAGCGC 2334
      :|||
260 .....ArgGly 261
2335 AATCTTAGTGCAGCGGAGACACCGACACTATAGCGTTACGCGCAAGCGCAC 2384
      :|||
262 AspAlaProAlaGlyGly.....AlaValProGlyGlyAlaVal 274
2385 CCAAAACGGC.....A 2395
      :|||
274 LProGlyGlyPheGlyProLeuLeuAspGlyTrpTyrGlyValAspValS 291
2396 ACCTCAGCGCTGTGGCAATGCCAGACACATTTAATCAAGCCACATTA 2445
      :|||
291 eAspSerThrValAspLeuAlaGlnSerIleValGlnAlaProGlnLeu 307
2446 AACGCGCAACACATCGCTTCGACAAATGCTTCATTTAATCTAAC... 2490
      :|||
308 GlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrValSerGlyG 324
2491 .....AACACGCGCGTACAAAACGGC..... 2511
324 YserLeuSerAlaProHisGlyAsnValIleGlnThrIleGlyAla 341
2512 .....AGTCAGCGCTTCGACACAC 2532
341 rArgPheProProAlaSerProLeuSerIleThrLeuGlnAlaGly 357
2533 GCTAAGCGCAAC.....GT 2546
      :|||
358 AlaArgAlaGlnGlyArgAlaLeuLeuTyrArgValLeuProGluProVa 374
2547 AAGCCATTCGCGACATCAACGGCAATGCTCCCTACCGCATTAAGCAGAT 2596
      :|||
374 LysLeuThrLeuAlaGlyGlyAlaGlnGlyGlnLysP.....IleV 389
2597 TCCATTTGAAACAGCGCTTACCGGAAAAAATGAGCGGCGGCAAGGAT 2646
      :|||
389 AlaIleThrGluLeuProProIleProGlyAlaSerSerGlyProLeuAsp 405
2647 ACGGCAATTAACCTTAAAGACAGCGAATGACGCTGCGTCGGCGACGGA 2696
      :|||
406 ValAlaLeu...AlaSerGlnAlaArgTrpThr...GlyAlaThrArgAl 420
2697 ATTAGCAATTTAAACCTTTGACAAGCGCACCATTTACATCAATTCGGCT 2746
      :|||
420 aValAspSerLeuSerIleAspAsnAlaThrTrpAlaMetThrAsp... 435
2747 ATCGACACGATCGGCGAGCGCGCAACCGGACGAGCGCGAGATGCGCG 2796
      :|||
436 .....AsnSerAsnValGlyAlaLeu 442
2797 CGCGCGCTGTGCGCGCGCTTCCTATTAATCGTTACCGCGCAACTTCGCG 2846
      :|||
443 ArgLeuAlaSerAspGlySer...ValAspPheGlnGlnProAlaGlnAl 458
2847 AGAATCCGCTTTCACACGCGTACGCTGAACGCGCAATTAAGCAAGTCAGG 2896
      :|||
458 a...GlyArgPheLysScyLysMetValAsp...ThrLeuAlaGlySerG 473
2897 GAACATTCGCGCTTATGTGCGAACTTCGCTACGCGAGCGGCAATTTG 2946
      :|||
473 LysLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeu 489

```



```

CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22201

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,269C
FILING DATE: 02-Jun-1995
ATTORNEY/AGENT INFORMATION:
NAME: Lebovitz, Richard M.
REGISTRATION NUMBER: 37,067
REFERENCE/DOCKET NUMBER: Popov-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 243-6410
TELEFAX: (703) 243-6333
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 910 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-460-269C-2

```

```

alignment_scores:
  Quality: 268.00      Length: 1156
  Ratio: 0.551        Gaps: 53
  Percent Similarity: 42.042  Percent Identity: 21.367

```

alignment_block:

US-09-303-518D-653 x US-08-460-269C-2 ..

Align seg 1/1 to: US-08-460-269C-2 from: 1 to: 910

```

1366 AAGGACCGCCTGTCCAAATCGGCAAGG.....ACGCT 1400
      ||| ||||| ||||| ||||| ||||| |||||
      2 AsnMetSerLeuSerArgIleValIysAlaIleProLeuArgIleThr 18
1401 GCTGTTCAAGCCCAAGG..... 1419
      ||| ||||| ||||| ||||| ||||| |||||
      18 rLeuAlaMetAlaLeuGlyAlaLeuGlyAlaIleProAlaIleHisAla 35
1420 ....GAAACCAAGGCTCGTCAGCGTGCGAGCGTAAGTCATCTTA 1464
      ||||| ||||| ||||| ||||| ||||| |||||
      35 sPTrpAsnAsnGlnSerIleValIysThrGlyIleuArgIleHisGlyIle 51
1465 GATCAGAGCGGCGATCAAGCAAAAAACAAGCCTTAGTCAATCGG 1514
      ||||| ||||| ||||| ||||| ||||| |||||
      52 HisIleGlnIleSerIleProGlyIleValArgIleThrAlaSerGlyThr 68
1515 CTGG....GTGAGCGGAGGAGGAGCGTGCACATGATGCCGATATCAGT 1561
      ||||| ||||| ||||| ||||| ||||| |||||
      68 rIleIysValSerGlyArgGlnAlaGlnIleGlyIleLeuLeuGlnAsn... 83
1562 TCACCCCGCAAACTGATTCGCTTCGCGGCGGAGCGATTGGATTGG 1611
      ||||| ||||| ||||| ||||| ||||| |||||
      84 .....ProAlaIleGlnIleuGlnPheArgIleGlnIleSerValIleThr 97
1612 AAGGGGATTCGCTTCGTCACCGCATTCGAAATACGATGACGAGGCG 1661
      ||||| ||||| ||||| ||||| ||||| |||||
      98 SerGly.....GlnLeuSerAspAspGlyIle 106
1662 GATGATGTCACACCAATCAAGACAAGATCCACGTTACCATTCAG 1711
      ||||| ||||| ||||| ||||| ||||| |||||
      106 e.....ArgArgPheLeuGlyIleThrValIleThrValIle 117
1712 GCATATAA.....GATATTACTACACCGGCAATACCAACAACCTGG 1752
      ||||| ||||| ||||| ||||| ||||| |||||

```

```

117 lAgIlyLeuValAlaAspHisAlaThrLeuAlaAsnValGlyAspThr 133
1753 GATACCAAAAAAGAAATTGCCACACAGCGTGTGGCGAGAAAGATGC 1802
      ||| ||||| ||||| ||||| ||||| |||||
      134 .....TrpAspAspGlyIleAl 140
1803 AACCAAAACGAAGCGCGCTCATATGCAATTCACACGGAAGAGCGG 1852
      ||||| ||||| ||||| ||||| ||||| |||||
      140 leuIyValAlaGlyGlu.....GlnAlaGlnIleSerIleAla 154
1853 ATCGCACTTACCTGCTTCGCGGCGAGCAAAATTA.....AACGCAT 1896
      ||||| ||||| ||||| ||||| ||||| |||||
      154 SerThrLeuGlnIleValIleGlyValGlnIleGlyArgGlyAlaAsn 170
1897 ATCAGCAACAAACGGAACACTGTTTTCAGCGGACGACGACACCGCA 1946
      ||||| ||||| ||||| ||||| ||||| |||||
      171 ValThrValGlnArgSerAlaIleValAspIleGlyLeu..... 183
1947 CGCCTACAAATCATTTAGAGACGGGCTGCTCA...AAATGGAAGTATCC 1993
      ||||| ||||| ||||| ||||| ||||| |||||
      184 .....HisIleGlyAlaLeuGlnSerLeuGlnProGlnAspLeuP 197
1994 CACAAGAGAAATCGTGTGGACAACGATGTGATCGACCGCACA..... 2037
      ||||| ||||| ||||| ||||| ||||| |||||
      197 rOProSerArgValValLeuArgAspThrAsnValIleThrAlaValProAla 213
2038 .....TTTAAAGC 2045
214 SerGlyAlaProAlaIleValSerValLeuGlyAlaSerGlyLeuThrLe 230
2046 GGAACACTTCATATTCAGGCGGACAGCGGCTGTTCGCGCAATGTTG 2095
      ||||| ||||| ||||| ||||| ||||| |||||
      230 uAspIleGlyHisIleThrGlyGlyArgAlaIle.....GlyAla 244
2096 CCAAGTGAAGCGGATTCGCAATTAAGCAATCAGCCCAACAGCTTTC 2145
      ||||| ||||| ||||| ||||| ||||| |||||
      244 la.....AlaMetGln 247
2146 GGTGTGCGACCGCATCAAGCCACACATCTGTACCTTGAGAC..... 2190
      ||||| ||||| ||||| ||||| ||||| |||||
      248 GlyAlaValIleHisLeuGlnArgAlaThrIleArgArgIleAspAlaLe 264
2190 ..... 2190
264 uAlaGlyIleAlaValIleProGlyIleAlaValIleProGlyIleAlaValIleProG 281
2191 .....TGACGCGGT 2199
281 IyGlyPheGlyProGlyIlePheGlyProValIleuAspGlyIleTrpIyGly 297
2200 CTGACAACTTGTACGAAAAAACCATTTACGACGATTAAGTATGCTTC 2249
      ||||| ||||| ||||| ||||| ||||| |||||
      298 Val..... 298
2250 ATTGACAAGACCGACATCAGAGGC...AATGTACGCTTCGCGATCAG 2296
      ||||| ||||| ||||| ||||| ||||| |||||
      299 .....AspValSerGlySerSerValGlnLeuAlaGln... 309
2297 CTCATTTAAATCTCACAGAGCTTGCACACATCAACGCGCAATCTTAAGTCA 2346
      ||||| ||||| ||||| ||||| ||||| |||||
      310 .....SerIleValGlnIleProGlyIleuGlyAlaIleIleArgVal 323
2347 GCGCGAGACAGCACTATACGTTACGCGGAC.....GCCACCCAAA 2390
      ||||| ||||| ||||| ||||| ||||| |||||
      324 GlyArgGlyAlaArgValIleThrValIleProGlyIleSerLeuSerIleProGln 340
2391 CGGCAACCTCAGCCTCGTGGCAATGCCCAAGCAACATTTATCAAGCC... 2439
      ||||| ||||| ||||| ||||| ||||| |||||
      340 sGlyAsnValIleGlyIleThrGlyIleAlaArgArgPheAlaIleProGlnIleAla 357
2440 .....ACATTAACGCAACACATGCGCTTCGAGCAATGCT 2475
      ||||| ||||| ||||| ||||| ||||| |||||
      357 lProLeuSerIleThrLeuGlnIleGlyAlaIleHisAlaGlnIleGlyAla 373

```



```

      805 ArgValArgAspIuGIyGlySerSerValIeuGIyArgIeuGIyIeuGI 821
4143 GGCAGATTATTCATCAACCGCGCACACATTTCCATCAGCCCTTATT 4192
      821 uValGIyArgIleGIleuGIyIleGIyArgIleGIyIleGIyIleGIyI 838
4193 TGAAGCTGTCTTATACCGATCCGCTTCGCGCAAAAGTCCGAACGGCGCTC 4242
      838 IeuYsAlaSerValIeuGIuGIuPheAspGIyAlaGIyThrValIleHstr 854
4243 AATACCCGCTATTGGCGCAGGATTTCGCGCAAAACCGCAGTGGCGAATG 4292
      855 AsnGIyIleAlaHisArgThrGIleuArgGIyThrArg...AlaGIyIle 870
4293 GGGCGTAAACGCCGAATCAAGGTTTCACGCTGTCCTCCACGCTGGCG 4342
      870 uGIyIle...LeuGIyMetAlaIleAlaIle 878
4343 CCGCCAAAGGCGCGCATTTGGAAGCG...CAGCACAGCGCGGCGATC 4386
      878 euGIyArgGIyHisSerIleuTyAlaSerTyGIyIleTySerIleGIyPro 894
4387 AATAGGCTACCGCTGG 4404
      895 LysLeuAlaMetProTirp 900

seq_name: /cgn2_6/plodata/1/laa/5A_COMB.pep:us-08-038-682-2
seq_documentation_block:
; Sequence 2, Application US/08038682
; Patent No. 5549897
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GENE III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Maltare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/038,682
; FILING DATE: 16-MAR-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 1038-293
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-038-682-2
alignment_scores:

```

```

      Quality: 265.00      Length: 1297
      Ratio: 0.438      Gaps: 73
      Percent Similarity: 46.646      Percent Identity: 21.126
alignment_block:
Align seg 1/1 to: US-08-038-682-2 from: 1 to: 1536
178 AATAAGGCAAGATTTCAGTCCGGGCAAGAAATATTGAGTTTACAACA 227
199 AsnHisGIyLeuIleThrValGIy...LysAspGIySerVal... 211
228 AAAAGGGAGATTGGTCGCAAAATGCATGACGAAGCCCGATGATGATT 277
212 ....AsnLeuIleGIyGIyValIleLysAsnGIy... 221
278 TTTCTGTGTATCGGTATACGCG...GTGGCGGCAATTGGCGGCGAT 321
222 ..GIyValIleSerValAsnGIyGIySerIleSerLeuIleAlaGIyGI 237
322 CA...TATATTGTGAG 335
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleTySerIleAl 254
336 CGTGGCACAATACGGCGGCTATACAAATGTTGTTGGTCGGAGGAA 385
254 AlaIleProGIuAsnGIuAlaValAsnLeuGIyAspIlePheAlaLysGIy 271
386 GCAT...CCGATTCACAGCAGCGTTTCT 411
271 LysIleAsnValAlaArgAlaIleThrIleArgAsnGIyGIyLysLeuSer 287
412 TTCCAATTTGCAAGAAATTAATTAACGACGACGCTAAGCGGCATCC 461
288 AlaAspSerValSerLysAsp...LysSerGIy... 297
462 TTATGCGCGCATATCATATGCGCGGTTTGCACAAATTTGTCACAGATG 511
298 .....AsnIleValIleuSer 303
512 CAGAACCTGTGAG...ATGACCACTTATATGAT 543
303 IalysGIuGIyGIuAlaGIuIleGIyGIyAlaIleSerAlaGIuAsnGIu 319
544 GGGTGAATACGCGCATTAATAATTAATACCTGATGCTGT...CG 587
320 GlnAlaLysGIyGIyLysLeuMetIleThrGIyAspLysValThrLeuGIy 336
588 AATCGAGCA...GGCAGCAATATT 610
336 sThrGIyAlaValIleAspLeuSerGIyLysGIuGIyGIyGIuThrTyIle 353
611 GCGCGTGTGATGA...GACGAACCAATAC... 639
353 euGIyGIyAspGIuArgGIyGIuGIyLysAsnGIyIleGIuIleuAlaLys 369
640 .....CGCGAAAGTTCATCATATGATGCAAGC... 666
370 LysThrSerLeuGIuLysGIySerThrIleAsnValSerGIyLysGIyLys 386
667 .....GCATATTCTTGG...CTGCGGTGGCAATA 694
386 sGIyGIyArgAlaIleValThrGIyAspIleAlaLeuIleAspGIyAsnI 403
695 CTTTGCACAATAATGATCAGGTGCGGCGACAGTCACTTAGTACGCA 744
403 IeAsnAlaGIu...GIySerGIy... 409
745 AATATTAAACATAGCCCATATGTTTATACCAAGACGAGGCTCATTTGG 794
410 .....AspIleAlaLysThrGIyGIyPheValGI 419

```


795 CGACAGTGGCTACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGT 844
|||||
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAAATATGGGATTTGCAACAGCAACCCCTATATAGAAAAAGCAAT 894
|||||
430AsnAlaIleValAspAla..... 435
895 GGCTTCACAGTAGTTCGAAGATTGGTCTTATGATGAAATCTTGTGCTGG 944
|||||
436LysGluThrLeuLeuAsp.....Ph 442
945 AGATACCATTCAGTATCTACAGACCATCAAAATGGGAAATACITTT 994
|||||
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerg 459
995 TTAACGACATATATATGCGCGAGCAAAATCATGCAAAATCAAAAC 1044
|||||
459 LuAspAspGluIleThrIleLysSerGly.....AsnSer 469
1045 TATCTCTACCTTATATGATTTAAACACAGAACGCTTCAATGTTTATGT 1094
|||||
470 AlaSerThrProLysArgAsnLysGluLysThrThr...LeuThrAsnTh 485
1095 TTCTTTATCCGAG.....ACAGCA 1114
|||||
485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAla 502
1115 GAGAACCTGTTAT..... 1128
502 sngIleArgIleIleValAsnSerSerIleAsnLeuSerAsnLysSerLeu 518
1129CATGCTGCAAGGTGGGCTCACAGTTATCGACC 1160
519 ThrLeuThrSerGluIleArgSerGlyGlyVal..... 530
1161 CAGACGCAATAT.....GGAGAAATA 1183
531GluIleAsnAsnAspIleThrThrGlyAspSphThrArgGlyAlaAsn 547
1184 TTTCCTTATGACAAAGAAAGTGAATGATGATCTTACGACGCAATC 1233
547 eThrIleIleThrSerGlyGlyThrValAsp.....ValHisLysAsnIle 561
1234 AACCAAGCGCGCGGCTTTGATTTTGAAGGTAATTTACGCTCGCC 1283
562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
1284 TAA.....AACACGAAACGTCGCAAGCG 1309
573 alysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGly 590
1310 CGGCGCTTATCATCAGTATGATGACGATACGCTTATGAAAGTAAAGCG 1359
590 lngIleThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605
1360 GTGGCAACGACGCGCTTCCAAATCGCAAGGCGCGCTGCTGCTCA 1409
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618
1410 AGCCAAAGGGGAAAC.....CAAGGCTCG 1435
618 rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyThr 635
1436 TCACGTCGGCGAGGTAAATCATCTTA.....GAT 1467
635 euAsnIle...SerGlyValAsnIleSerMetValLeuProLysAsn 650
1468 CAGACGCGGACGATCAAGCAAAACAAAGCCTT..... 1503
651 GluSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThr 667
1504AGTGAATCGGC.....TTGGTCAGCGGCAAGG 1531

667 rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArg 684
1532 GCAGC.....GNG 1539
684 LysSerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle 700
1540 CAACGTAAAGCCGATATATCATGTCACCCGCAAA..... 1575
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValas 717
1576CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsn 734
1580 ATTTCGCTTCCGCGCGGACGCTTGGATTTGAACGCGCATTCGCTTGG 1629
734 rAlaSerPheAsnGlyAsnIleSerValSerGlyGlySerValAsp 750
1630 TTCACCGCAT.....CAAAATCCGATGAAGGCGGATGATGT 1670
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValAla 767
1671 CAACCAAT.....CACACAAAG 1690
767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuArgPheLys 784
1691 AATCCACGCTTACCATTCAGCG.....ATAAGATATTA 1728
784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800
1729ACAAACCGCAATAC..... 1743
801 AsnAlaThrArgGlyAlaAsnIleThrLeuLeuGlnValGluGlyThrAsp 817
1744AACACTTGGATACCAAAAGAAATGCCACAGCGT. 1782
817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGly 834
1783TGCTTGGCGAGAAAGATCA...ACCAAAACGACGAGCGG 1821
834 LysnIleThrPheGlySerArgLysAlaValThrGluIleGluLys 850
1822 CTCATTCGATTTACCAACCGGACGACGATCGCACTTA..... 1863
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGly 864
1863 1863
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIle 881
1864CTGCTTCCGCGCGA...ACAAATTTAAACGCG 1893
881 leAsnSerGlyAsnLeuThrAlaGlyLysAlaIleValAsnIleAlaGly 897
1894 AATATCAG...CAACAAACGCGCAACTGTTTTCACGCGA..... 1933
898 AsnLeuThrValGluSerAsnAlaAsn...PheLysAlaIleThrAsn 913
1934GACGACACCGCGCGCTCAATC 1957
913 heThrPheAsnValGlyLysLeuPheAspAsnLysGlyAsnSerAsnIle 929
1958 ATTAGAGCGGTGTCAAAATGAGATATCCACAGCAAGAAATC 2007
930 SerIleAlaLysGlyLysAlaArgPheLysAspIleAspAsnSerLysAs 946
2008 GTGTGGACACGATTCATCGACCGC.ACATTTAAAGCGGAAACTTCC 2056
946 nLeuSerIleThrThrAsnSerSerThrTyrArg..... 958
2057 ATATTGAGGCGGACAAAGCGGTGTTCCCGCAATGTTCCAAAGTGA 2106

```

2902 . TTCCGCTTATGTGGACACTTGCGCTACCGCAGGCCCAATTGAA 2948
      ::::| | | | | | | | | | | | | | | | | | | | | | | |
1243 GtUalYlsAlatHrGlUsErLeuHrThrGlNsErnsErLySIlEly 1259
      : : : : : : : : : : : : : : : : : : : : : : : :
2949 GCTGCAGGAAGATTCCGAGGCACTTACCCTTGCGTCG..... 2988
      || | | | | | | | | | | | | | | | | | | | | | |
1259 S...AlatHrThrGlyGuaIaaNsValThrSerAlathrcLylHrIEg 1275
      - - - - - - - - - - - - - - - - - - - - - - - -
2989 .....AACATTAACCGGCAACGACCCGTAAGCTCTGAGCAA 3024
      : : : : : | | | | | | | | | | | | | | | | | |
1275 LyglYlHrIIeSerGlyAsnHrValasNvalHrrAaAsAlaGlYaSp 1291
      : : : : : : : : : : : : : : : : : : : : : : : :
3025 TTGACGGTA.....GTGGAGAAGAAAGCAA 3050
      || | | | | | | | | | | | | | | | | | | | | | |
1292 LeuHrValGlYasnGLyagIUleAsNAlatrHrLuGlYalaIaIaTh 1308
      : : : : : : : : : : : : : : : : : : : : : : : :
3051 CACACCGCGCTGCCGAAATCTTAATTTACCCTTCGCAAAAGMACGCTGC 3100
      : : : : : | | | | | | | | | | | | | | | | | |
1308 rLeuHrThrSeSrErLyLSLeuHrThrGUalSaSeSrItler 1325
      : : : : : | | | | | | | | | | | | | | | | | |
3101 ATCCGCGCATGAGCGTATCAGTTATTCGCAAAAGCGCGAGTTCCGC 3150
      : : : | | | | | : : : : : | | | | | | | | | |
1325 hrerAlaYslYelInVaIsNrLeuSerAlGIlnspISySeVAlAI 1341
      : : : : : : : : : : : : : : : : : : : : : : : :
3151 CTCGCAATPACCGGTCAAAACAAGACGTTCCGACA 3187
      | | | | | : : : : : | | | | | | | | | | | |
1341 acLYSerIEaSnAlaLaSNvalThrLeuAsnThr 1353

```

```

seq_name: /cgn2_6/prodata/1/aa/5A_COMB.pep:us-08-302-832-2
seq_documentation_block:
? Sequence 2, Application US/08302832
? Patent No. 5603938
?
? GENERAL INFORMATION:
?   APPLICANT: Barenkamp, Stephen J
?   TITLE OF INVENTION: High Molecular Weight Surface Proteins
?   TITLE OF INVENTION: of No. 5603938-Typeable Haemophilus
?   NUMBER OF SEQUENCES: 8
?   CORRESPONDENCE ADDRESS:
?     ADDRESSEE: Shoemaker and Mattare, Ltd.
?     STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
?     STREET: Bldg. 1
?     CITY: Arlington
?     STATE: Virginia
?     COUNTRY: U.S.A.
?     ZIP: 22202-0286
?
? COMPUTER READABLE FORM:
?   MEDIUM TYPE: Floppy disk
?   COMPUTER: IBM PC compatible
?   OPERATING SYSTEM: PC-DOS/MS-DOS
?   SOFTWARE: PatentIn Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
?   APPLICATION NUMBER: US/08/302,832
?   FILING DATE: 16-SEP-1994
?   CLASSIFICATION: 435
?
? PRIOR APPLICATION DATA:
?   APPLICATION NUMBER: GB 9205704.1
?   FILING DATE: 16-MAR-1992
?
? PRIOR APPLICATION DATA:
?   APPLICATION NUMBER: US pct/us93/02166
?   FILING DATE: 16-MAR-1993
?
? ATTORNEY/AGENT INFORMATION:
?   NAME: Berkstresser, Jerry W
?   REGISTRATION NUMBER: 22,651
?   REFERENCE/DOCKET NUMBER: 1038-404
?
? TELECOMMUNICATION INFORMATION:
?   TELEPHONE: (703) 415-0810
?   TELEFAX: (703) 415-0813
?
? INFORMATION FOR SEQ ID NO: 2:
?   SEQUENCE CHARACTERISTICS:
?     LENGTH: 1536 amino acids
?     TYPE: amino acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-302-832-2

```

```

alignment_scores:
  Quality: 265.00      Length: 1297
  Ratio: 0.438        Gaps: 73
  Percent Similarity: 46.646  Percent Identity: 21.126

```

```
alignment_block:
```

```
US-09-303-518d-653 x US-08-302-832-2 ..
```

```
Align seg 1/1 to: US-08-302-832-2 from: 1 to: 1536
```

```

178 AATAAGGAGAGTTGACGCGGGGGAAGATTTAGGTTTACAA 227
    ||| ||| :|||:||||| ||||| :|||
199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal... 211
    :|||:||||| :|||:||||| :|||:|||||
228 AAAAGGAGAGTTGCGCAATCGATGACGAAGCCCGATGATTGATT 277
    :|||:||||| :|||:||||| :|||:|||||
212 .....AsnLeuIleGlyGlyValValLysAsnGlu..... 221
    :|||:||||| :|||:||||| :|||:|||||
278 TTCTGTGCTATCGCTAACGGC.....GTGGCGCATTTGGCGGCGAT 321
    :|||:||||| :|||:||||| :|||:|||||
222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuIleGlyGln 237
    :|||:||||| :|||:||||| :|||:|||||
322 CAA.....TATATTGTGTGAG 335
    ||| ||| :|||
238 LysIleThrIleSerAspIleIleAsnProIleThrIleThrSerIleAl 254
    :|||:||||| :|||:||||| :|||:|||||
336 CGTGGCACAATACGGCGCTATACAAATGTTGTTGCGGGAGGAA 385
    :|||:||||| :|||:||||| :|||:|||||
254 aAlaProGluAsnGluIleValAsnLeuGlyAspIlePheAlaLysGly 271
    :|||:||||| :|||:||||| :|||:|||||
386 GCAT.....CCGATCAGCAGCGCTTTCT 411
    :|||:||||| :|||:||||| :|||:|||||
271 LysAsnIleAsnValArgAlaIleThrIleArgAsnGlnGlyLysLeuSer 287
    :|||:||||| :|||:||||| :|||:|||||
412 TACCAATTTGAAAGAAATAATTTAAGACGAGACTAAGCCCATCC 461
    :|||:||||| :|||:||||| :|||:|||||
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
    :|||:||||| :|||:||||| :|||:|||||
462 TTATGGCGCGCATTTATCATATCCGCGCTTGCCAAATTTGTCACAGAT 511
    :|||:||||| :|||:||||| :|||:|||||
298 .....AsnIleValLeuSerA 303
    :|||:||||| :|||:||||| :|||:|||||
512 CAGAACCTGTTGAG.....ATGACCAGTTTATGAT 543
    :|||:||||| :|||:||||| :|||:|||||
303 LysGlyGluGlyIleValIleGlyGlyValIleSerAlaGlnAsnGln 319
    :|||:||||| :|||:||||| :|||:|||||
544 GGGTGAATATACGCTGATTAAATAATACCCGATGCTGT.....CG 587
    :|||:||||| :|||:||||| :|||:|||||
320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLys 336
    :|||:||||| :|||:||||| :|||:|||||
588 AATCGAGACA.....GGCAGACAAATATT 610
    :|||:||||| :|||:||||| :|||:|||||
611 GCGGCTGATGAA...GACGACCCCAATAAC..... 639
    :|||:||||| :|||:||||| :|||:|||||
353 euGlyGlyAspGluArgGlyGlyLysAsnGlyIleGlnLeuAlaLys 369
    :|||:||||| :|||:||||| :|||:|||||
640 .....CGGAAAGTTCATATCATATTGCAAC..... 666
    :|||:||||| :|||:||||| :|||:|||||
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGlyLys 386
    :|||:||||| :|||:||||| :|||:|||||
667 .....GCATATCTTGG.....CTGCTGCGTGACATA 694
    :|||:||||| :|||:||||| :|||:|||||
386 serGlyArgAlaIleValIleThrPcLysAspIleAlaLeuIleAspLysAsnI 403
    :|||:||||| :|||:||||| :|||:|||||
695 CCTTTGCACAAATGATTCAGGTGTGGCAGCAGTCACTAGTAGGCA 744
    :|||:||||| :|||:||||| :|||:|||||

```

```

403 LeuAsnAlaGln...GlySerGly..... 409
    ||||| |||||
745 AATAATTAACATAGCCCATATGCTTTTACCACAGAGCGTCATTTGG 794
    :|||:||||| :|||:||||| :|||:|||||
410 .....AspIleAlaLysThrGlyGlyPheValG 419
    :|||:||||| :|||:||||| :|||:|||||
795 CGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAGTGT 844
    :|||:||||| :|||:||||| :|||:|||||
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
    :|||:||||| :|||:||||| :|||:|||||
845 TAATTAATGGGTATTCGAAACAGGCAACCCCTATATAGAAAAAGCAAT 894
    :|||:||||| :|||:||||| :|||:|||||
430 .....AsnAlaIleValAspAla..... 435
    :|||:||||| :|||:||||| :|||:|||||
895 GCGCTTCACACTAGTTCGTAAAGATGTTTATGATGAATCTTTGCTGG 944
    :|||:||||| :|||:||||| :|||:|||||
436 .....LysGluThrPheLeuAsp.....Phe 442
    :|||:||||| :|||:||||| :|||:|||||
945 AGATACCATTCAGTATTCACGAACACATCAAAATGGAATACCTTTT 994
    :|||:||||| :|||:||||| :|||:|||||
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459
    :|||:||||| :|||:||||| :|||:|||||
995 TTACGACAAATTAATATGCGCGACGAAATAATGATGCCAACTAAACAC 1044
    :|||:||||| :|||:||||| :|||:|||||
459 LysAspAspGluThrThrGlySerGly.....AsnSer 469
    :|||:||||| :|||:||||| :|||:|||||
1045 TATTCCTACCTATATAGATAAAACACGACACCGTTCAATGTTTAATGT 1094
    :|||:||||| :|||:||||| :|||:|||||
470 AlaSerThrProLysArgAsnLysGlyLysThrThr..LeuThrAsnTh 485
    :|||:||||| :|||:||||| :|||:|||||
1095 TTCTTATTCGAG.....ACAGCA 1114
    :|||:||||| :|||:||||| :|||:|||||
485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAla 502
    :|||:||||| :|||:||||| :|||:|||||
1115 GAGACCTGTTTAT..... 1128
    :|||:||||| :|||:||||| :|||:|||||
502 snGlnArgIleThrValAsnSerSerIleAsnLeuSerAsnLysSerLeu 518
    :|||:||||| :|||:||||| :|||:|||||
1129 .....CATGCTGACGTGGCGGTCAACAGATTATGACGC 1160
    :|||:||||| :|||:||||| :|||:|||||
519 ThrLeuThrSerGluGlyArgSerGlyGlyVal..... 530
    :|||:||||| :|||:||||| :|||:|||||
1161 CAGACTGAATAT.....GAGAAATA 1183
    :|||:||||| :|||:||||| :|||:|||||
531 GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsn 547
    :|||:||||| :|||:||||| :|||:|||||
1184 TTTCCTTATTCGACAAAGAAAGTGAATGATCTTACCAACACATC 1233
    :|||:||||| :|||:||||| :|||:|||||
547 euThrIleThrSerGlyGlyThrValAsp.....ValHisLysAsnIle 561
    :|||:||||| :|||:||||| :|||:|||||
1234 AACCAAGCGCGGCGGTGTGATTTGAGGGAATTTTACGCTCCGCC 1283
    :|||:||||| :|||:||||| :|||:|||||
562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrLeu 573
    :|||:||||| :|||:||||| :|||:|||||
1284 TAA.....AACACGAAACGTGCAAGGCG 1309
    :|||:||||| :|||:||||| :|||:|||||
573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590
    :|||:||||| :|||:||||| :|||:|||||
1310 CGGGCGTTCATATCAGTATGCGACGATCCGTTACTTGAAGAAATTAACGC 1359
    :|||:||||| :|||:||||| :|||:|||||
590 lnglyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605
    :|||:||||| :|||:||||| :|||:|||||
1360 GTGGCAACGACCGCTCTCCAAATCGCGCAAGCAGCGCTGCGGTCA 1409
    :|||:||||| :|||:||||| :|||:|||||
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618
    :|||:||||| :|||:||||| :|||:|||||
1410 AGCCAAAGGGGAAAC.....CAAGCTCGG 1435
    :|||:||||| :|||:||||| :|||:|||||
618 rThrLysArgThrAsnLysThrAlaIleThrAsnLysPheGluLysThr 635
    :|||:||||| :|||:||||| :|||:|||||
1436 TCAGCGTGGCGACGATTAAGTCATCTTA.....GAT 1467
    :|||:||||| :|||:||||| :|||:|||||

```

```

635 euhsnlle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGCGCGACGATCAAGCAGCAAAAAAGAGCTTT..... 1503
651 GluSerGlyTyrAspLysPheLysGlyArgThrTyrTrpAsnLeuThrSe 667
1504 .....ACTGAATCGGC.....TTGGTCAGCGCGCAGG 1531
667 rLeuAsnValSerGluSerGlyLupheAsnLeuThrIleAspSerArg 684
1532 GGAGC.....GTG 1539
684 LysSerSperAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyLe 700
1540 CAACGTAAATGCGGATATCATGTTCAACCCCGACAAA..... 1575
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
1576 .....CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsn 734
1580 ATTTCGGCTTTGCGCGCGACGCTTTGATTTGAACGGCGCATTTGCTTTG 1629
734 yAlaSerPheAsnGlyAsnIleSerValSerGlyGlySerValAsp 750
1630 TTCCACCGCAT.....CAAAATACGATGAGGGCGCATGTGT 1670
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValIle 767
1671 CACCCACAT.....CACACAAAG 1690
767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuThrPheLys 784
1691 AATCCACCGTTACCTTACAGCG.....ATAAAGATATTACT... 1728
784 hrSerGlySerThrLysThrGlyPheSerIleGlyLysAspLeuThrLeu 800
1729 ...ACACCCGGCAATTAAC..... 1743
801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGlnGlyThrAspG 817
1744 .....AACAACCTTGATACCAAAAAAATAATGCTTACAAACGCT 1782
817 yMetIleGlySerGlyLeuAlaLysAsnIleThrPheGlnGlyG 834
1783 .....TGCTTGGCGAGAAAGATGCA...ACCAAAACGAGGGCGG 1821
834 LysAsnIleThrPheGlySerArgLysAlaValThrGlnIleGlnGlyAsn 850
1822 CTCATCTGATTTACCAACCGGAAGACGATGCACTTA..... 1863
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864
1863 ..... 1863
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIle 881
1864 .....CTGCTTCCGGCGGA.....ACAATTTAAACGGC 1893
881 leAsnSerGlyAsnLeuThrAlaGlyLysnIleValAsnIleAlaGly 897
1894 AATATACAG...CAACAAACGCGCAACGTGTTTTCAGCGGCA..... 1933
898 AsnLeuThrValGluSerAsn..AlaAsn...PheLysAlaIleThrAsp 913
1934 .....GACCGACACCGCAGCGCTTCAATC 1957
913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyLysSerAsnIle 929
1958 ATTTCAGAGCGGTGTCAAAATGAAAGTATGCCACAGAGAAATC 2007
930 SerIleAlaLysGlyLysAlaThrPheLysAspIleAspAsnSerLysAs 946
2008 GTGTGGACACAGCATTTGATCGACCGC..ACATTTAAACGGAACCTTCC 2056
946 nLeuSerIleThrThrAsnSerSerThrTyrArg..... 958
2057 ATATTACAGGGCGGACACAGCGGTGTTTCCCGCATGTTGCCAAAGTGGA 2106
959 .....ThrIleIleSerGlyAsnIleThrAsnLysAsn 969
2107 GCGGATTCGATTTAAACATCAGCGC.....CAAGCATTTT 2144
970 GlyAspLeuAsnIleThrAsnGlnGlySerAspThrGlnMetGlnIleG 986
2145 CGGTGCGACCGCATCAAGACACACATGTGACAGCTTCGAGCTGA 2194
986 yGlyAspValSerGlnLysGlnGlyAsnLeuThrIleSerSerAspLys 1003
2195 CGGTCTGACAGTTGTACCGAAAAAACCATTACC...GACGATAAAGTG 2241
1003 leAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015
2242 ATTGCTTCATTGACGACAGACGACATCAGACGATGTCAGCTTGGCGA 2291
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThr 1032
2292 TCACGT.....CATTTAATCTCAGAGACTT 2319
1032 eLysThrLysGlnLeuLysLeuThrGlnAspLeuAsnIleSerGlyPhe 1049
2320 .....GCCACATCAACGGC 2334
1049 snLysAlaGlnIleThrAlaLysAspLysSerAspLeuThrIleGlyVal 1065
2335 AATCTTACTGACGCGGACGACACGACATTAACGCTTAACGCGCAC 2384
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGln 1082
2385 CCAAAACGCGCAACCTCAGCTCGGGCAGTGGCCCAAGCAACTTAATC 2434
1082 LysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSer 1099
2435 AAGCCACATTTAAACGCGACACATCGGCTTCGACATGCTTCATTTAAT 2484
1099 yValGlnThrSerGlySerAsnAsnAsnThrGlnLysPheSerSerAsp 1115
2485 CTAGCAACAACGCGGTACAAACGCGACGTGACGCTTCCGACAAACG 2534
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAsn 1132
2535 TAAAGCAAC.....GTAAAGCATTCGCGACATCAAGCGCATGTGTC 2574
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGlnIleThr 1149
2575 .....TCCCTAAGCCGATTAAGCGAGTATTCATTTGAAAAACAG 2613
1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGlnIleThr 1165
2614 CGCTTACCGGAAAAATCAAGCGCGCGCAAGGATAG.....GCAT 2654
1166 AlaGlnThrGlySerIleLeuGlnGlyIleGluSerSerGlySerVal 1182
2655 ACCTTTAAAAAGACAGGATGAGCGCTGCGGCGACAGGAAATTAGGCA 2704
1182 ThrLeuThrAlaThrGlnGlyAlaLeu.....AlaValSerA 1195
2705 ATTTCACCTTGACAAACGCGACATTAACATCAATTCGCGCTATCGACAC 2754
1195 snIleSerLysThrValThrValThrAlaAsnSer..... 1207
2755 GATGCGGCGCGCGCAACCGCGAGTGCAGGACAGATGCGCGCGCGCG 2804
1208 .....GlyAlaLeuThrThrLeuAlaGlySer..... 1216

```

```

2805 TTGGCGCCGCTTCCCTATTATCCGTACGCCGCCAACCCTGCGAGCAATCC 2854
      :      :      :      :      :      :      :      :      :
1217 .ThrlleYgLyThrglnserValthr.....ThrserserGlnserG 1231
      :      :      :      :      :      :      :      :      :
2855 GTTTCACACGCTGACGTAACGCAAAATTGACGCTGACGAGACA... 2901
      :      :      :      :      :      :      :      :      :
1231 Ly.....AspIleGlyLyThrlleSerGly...GlyThrVal 1242
      :      :      :      :      :      :      :      :      :
2902 ...TTCCGCTTTATGTCGAACTCTGCTGCTACCGAGCGGCAAAATTGAA 2948
      :      :      :      :      :      :      :      :      :
1243 GluVallysalathrglnserleuthrThrglnserAsnserlyIlely 1259
      :      :      :      :      :      :      :      :      :
2949 GCTGGCGAAAGTTCGAGGCACTTACACCTTGCTGTC..... 2988
      :      :      :      :      :      :      :      :      :
1259 s...AlathrThrglyGlnValasnValThrserAlathrglyThrIleG 1275
      :      :      :      :      :      :      :      :      :
2989 .....ACAAATRCGCGCAACGACACCCGTAATCTCGAGCAA 3024
      :      :      :      :      :      :      :      :      :
1275 lyGlyThrlleSerGlyAsnThrValasnValThrAlaasnAlaGlyasp 1291
      :      :      :      :      :      :      :      :      :
3025 TTGACGGTA.....GTGAGAGAAAGACAA 3050
      :      :      :      :      :      :      :      :      :
1292 LeuthrValIglYasnGlyAlaGlnIleasnAlathrglnGlyAlaIath 1308
      :      :      :      :      :      :      :      :      :
3051 CACACCGCTGTCGAAATTTTAACTTACCCCTGCAAAACGACACGCTG 3100
      :      :      :      :      :      :      :      :      :
1308 rleuthrThrserserGlyLysleuthrThrglnIleSerSerHisIle 1325
      :      :      :      :      :      :      :      :      :
3101 ATCGCGGGGCAATGGGCTATACGCTTATCCGCAAGACGGGAGTTCG 3150
      :      :      :      :      :      :      :      :      :
1325 hrSer..AlaLysGlyGlnValasnLeuSerAlaGlnAspGlySerVal 1341
      :      :      :      :      :      :      :      :      :
3151 CTGCATATCCGCTCAAGAACACAGCTTCCGACA 3187
      :      :      :      :      :      :      :      :      :
1341 eGlySerIleAsnAlaIaAsnValThrLeuAsnThr 1353
      :      :      :      :      :      :      :      :      :
seq_name: /cgn2_6/ptodata/1/1aa/5b_COMB.pep:us-08-530-198-2
seq_documentation_block:
; Sequence 2, Application US/08530198
; Patent No. 5869065
; GENERAL INFORMATION:
; APPLICANT: BARENKAMP, STEPHEN J
; APPLICANT: ST. GEME III, JOSEPH W
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT SURFACE PROTEINS
; TITLE OF INVENTION: OF NON-TYPEABLE HAEMOPHILUS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; STREET: Bldg. 1
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/530,198
; FILING DATE: 13-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: BERKSTRESSER, JERRY W
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: JWB-1186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 415-0813
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1536 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-530-198-2

alignment_scores:
      Quality: 265.00      Length: 1297
      Ratio: 0.438      Gaps: 73
      Percent Similarity: 46.646      Percent Identity: 21.126

alignment_block:
US-09-303-518D-653 x US-08-530-198-2 ..

Align seg 1/1 to: US-08-530-198-2 from: 1 to: 1536

178 AATAAGGCAAGTTTGCATGCGGGCGAAGATTTGAGGTTTCAACAA 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
199 AsnHsGlyLeuIleThrValGly..LysAspGlySerVal..... 211
228 AAAAGGGAGTTGTCGCGCAATCGATGACGAAAGCCCGCATGATTGTT 277
    : : : : : : : : : : : : : : : : : : : : : :
212 .....AsnleuIleGlyLysValLysAsnGlu..... 221
278 TTTCGTGCTATCGCTACAGGC.....GTGGCGCATTCGGCGCGAT 321
    : : : : : : : : : : : : : : : : : : : : : :
222 ..GlyValIleSerValasnGlyLysSerIleSerLeuIleAlaGlyGln 237
322 CNA.....TATATTGTCAG 335
    : : : : : : : : : : : : : : : : : : : : : :
238 LysIleThrlleSerAspIleIleAsnProThrlleThrIleSerIleAl 254
336 CCGTGCACATACGCGGCTATACCAATGTGATTGGTGGAGAGGAA 385
    : : : : : : : : : : : : : : : : : : : : : :
254 aAlaProGlnAsnGlnAlaValasnLeuGlyAspIlePheAlaLysGly 271
386 GCAAT.....CCGATCAGACACCGCTTTTCT 411
    : : : : : : : : : : : : : : : : : : : : : :
271 LysIleAsnValarGlnAlaIathrlleArGasnGlnGlyLysLeuSer 287
412 TACCAATTTGTCAAAAGAAATATATTAATTAACGAGGACTACGGCCATCC 461
    : : : : : : : : : : : : : : : : : : : : : :
288 AlaAspSerValSerLysasp.....LysSerGly..... 297
462 TTATGCGCGCATTTATCATATCCGCGCTTGCACAAATTTGTCACAGATG 511
    : : : : : : : : : : : : : : : : : : : : : :
298 .....AsnIleValleuSerA 303
512 CAGAACCTGTGAG.....ATGACCAAGTTATATGAT 543
    : : : : : : : : : : : : : : : : : : : : : :
303 LalsGlnGlyGlnAlaGlnIleGlyGlyValIleSerAlaGlnAsnGln 319
544 GGGTGAATATACCGATTTAATAATACCTGATCGGT.....CG 587
    : : : : : : : : : : : : : : : : : : : : : :
320 GlnAlaLysGlyLysLysLeuMetIleThrglyAspLysValThrLeuLy 336
588 AATCGAGACA.....GGCAGACAATATT 610
    : : : : : : : : : : : : : : : : : : : : : :
336 sthnglyAlaValIleAspLeuSerGlyLysGlnGlyGlnThrTrpYL 353
611 GCGGCTGTGATGA...GACGAACCAATAC..... 639
    : : : : : : : : : : : : : : : : : : : : : :
353 euGlyGlyAspGlnArGlyGlyGlyLysAsnGlyIleGlnIleuAlaLys 369
640 .....CGCGAAGTTCATATCATATTGCAAGC..... 666
    : : : : : : : : : : : : : : : : : : : : : :
370 LysThrSerLeuGlnLysGlySerThrIleAsnValSerGlyLysGlnLy 386
667 .....GCATATTCTTGG.....CTCGCGGTGGCAATA 694
    : : : : : : : : : : : : : : : : : : : : : :
386 sGlyGlyArGlnAlaIleValThrGlyAspIleAlaLeuIleAspGlyAsnI 403

```

```

695 CCTTGCACAAAATGATCAGGTGGTGGACAGTCACTTAGTAGCGAA 744
    ||||| |||||
403 leasnaIaGln...GlySerGly..... 409
745 AAAATTAACATAGCCCATATGTTTTCACCAACAGAGCGTCATTGG 794
    |||||
410 .....AspIleAlaIysThrGlyGlyIrpheValG 419
795 CGACAGTGGCTCACCAGTGTTCATTCATGATGCCAAAAGCAAAAGTGT 844
    |||||
419 uThrSerGlyHisAspIeuPhelIeuYasp..... 429
845 TAAATTAATGGGTATTCGCAACAGCGCAACCCCTATATAGAAAAAGCAAT 894
    |||||
430 .....AsnAlaIleValAspAla..... 435
895 GCGTCCAGCTAGTTCGTAAGATGTGTCATAGTAAGAAATCTTGCTGG 944
    |||||
436 .....LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCCATTCAATATTCACAGCACATCAAAATGGCAATACTTTT 994
    |||||
442 eaSpAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerg 459
995 TTAAACGCAATATTAATGGCGCAGGAAAAATGATGCCAAACATAAGAC 1044
    |||||
459 IuAspAspGluTrpThrGlySerGly.....AsnSer 469
1045 TATTCCTCAGCTTATAGATTAAACAGACAGCTTCATTTGTTAAAGT 1094
    |||||
470 AlaSerThrProLysArgAsnLysGluYsrThr...LeuThrAsnThr 485
1095 TTCTTTATCCGAG.....ACAGCA 1114
    |||||
485 rThrIeuGluSerIleLeuLysGlyThrPheValAsnIleThrAla 502
1115 GAGAACCTGTTAT..... 1128
    |||||
502 snGlnArgIleThrValAsnSerSerIleAsnLeuSerAsnGlySerIeu 518
1129 .....CATGCTCAGAGTGGGTCAACAGTATTCGAC 1160
    |||||
519 ThrLeuTrpSerGluGlyArgSerGlyGlyVal..... 530
1161 CAGACGGAATAT.....CGAAGAAATA 1183
    |||||
531 .GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547
1184 TTTCCTTATTCGCAAGAAAGTGAATGATACCTTACACAGACATC 1233
    |||||
547 euThrIleThrSerGlyGlyTrpValAsp.....ValHisLysAsnIle 561
1234 AACCAAGCGCGCGGCGGTTTGTATTTGAGGTAAATTTACGCTCGCC 1283
    |||||
562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
1284 TAAA.....AACACGGAACGTGGCAGCGC 1309
    |||||
573 alyGlnAlaSpIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590
1310 CGGCGCTCATATCAGTATGACAGTACCGTTCATTCGGAAGTAAACGGC 1359
    |||||
590 lngIlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605
1360 GTGCAACACGACCGCGCTGCCAAATGCGCAAGACAGCGCTGCTGTTCA 1409
    |||||
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618
1410 AGCCAAAAGGGAAC.....CAAGGCTCGG 1435
    |||||
618 rThrLysArgThrAsnLysThrAlaIleThrAsnLysPheGluGlyThrL 635

```

```

1436 TCACGCTGGCGGACGATAAGTCATCTTA.....GAT 1467
    |||||
635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGCGCGGACATCAAGGCAAAAAACAAGCTTT..... 1503
    |||||
651 GluSerGlyTyrrAspLysPheLysGlyArgThrTyrrAsnLeuThrSe 667
1504 .....AGTCAATCGGC.....TGGTCAGCGGCGAGG 1531
    |||||
667 rLeuAsnValSerGluSerGlyLurPheAsnLeuThrIleAspSerArgG 684
1532 GACAG.....GTG 1539
    |||||
684 LysSerAspSerAlaGlyThrLeuThrGlnProTyrrAsnLeuAsnGlyLe 700
1540 CAACGTGAATCCGATATCATGTTCAACCCCGACAA..... 1575
    |||||
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
1576 .....CCT 1579
717 nPheAspIleLysAlaProIleGlyLleAsnLysTyrrSerIeuAsnT 734
1580 ATTTGGGCTTTCGGCGGACGCTTGGATTGACGGGCATTCGCTTTCG 1629
    |||||
734 rYrLaserPheAsnGlyAsnIleSerValSerGlyGlySerValAsp 750
1630 TTCACCGGATT.....CAAAATCCGATGAAGGGGAGGATGTTGT 1670
    |||||
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValIle 767
1671 CAACCAAT.....CAAGCAAG 1690
    |||||
767 eaSnSerLysTyrrPheAsnValSerThrGlySerSerLeuArgPheLysT 784
1691 AATCCACGCTTACATTCACAGC.....AATAAGATATTACT... 1728
    |||||
784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800
1729 ..ACACCGGCAATAC..... 1743
    |||||
801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspG 817
1744 .....AACAACTTGATACCAAAAAAGAAATTCCTACACGCT. 1782
    |||||
817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyG 834
1783 .....TGGTTTGGCGAGAAAGATGCA...ACCAAAACGACGCGCG 1821
    |||||
834 LysAsnIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsn 850
1822 CTCATTCGATTTACCAACCGGAAGAGCGGATCGCACTTTA..... 1863
    |||||
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864
1863 ..... 1863
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleI 881
1864 .....CTGCTTTCGCGCGA.....ACAAATTTAAACGCG 1893
    |||||
881 leAsnSerGlyAsnLeuThrAlaGlyLysAsnIleValAsnIleAlaGly 897
1894 AATATCAG...CAAAACAAAGGCAAAAGTTCCTTTCACGCGCA..... 1933
    |||||
898 AsnLeuThrValGluSerAsnAlaAsn...PheLysAlaIleThrAsp 913
1934 .....GACCGACACGCGACGCTCAATC 1957
    |||||
913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929
1958 ATTTAGGAACGGGTGTCAAAAATGGAAGTATCCACACAGGAAATC 2007

```

```

930 SerIleAlaIalysglIylAlaArgPheIylAspIleAspAsnSerIylSas 946
2008 GTGGGACAAACGATGTGATCGACCGC..ACATTTAAAGCGGAAACTTCC 2056
946 nLeuSerIleThrAsnSerSerThrTyrArg..... 958
2057 ATATTACAGCGGACAAAGCGGTGTTCCCGCAATGTTCGCAAGTGAA 2106
959 .....ThrIleSerGlyAsnIleThrAsnIylAsn 969
2107 GCGCATTTGGCATTTAAGCATACGCC.....CAGACAGTTT 2144
970 GlyAspLeuAsnIleThrAsnIylGlySerAspThrGluMetGluIleG1 986
2145 GGGTGTGCGACCGCATCAAGCCACACATCTGACAGTTGCGGACTGA 2194
986 yGIylAspValSerGlnIylGlyAsnLeuThrIleSerSerAspIylS1 1003
2195 CGGCTGACAGTTGTACGAAACCATTTACC...GACGATAAAGTG 2241
1003 leAsnIleThr.....LysGluIleThrIleLysAlaGlyVal 1015
2242 ATTGCTTATGACAGACGACATGACAGCAATGTGACCTTGCCGA 2291
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAlaAsnLeuThrI1 1032
2292 TCACGCT.....CATTTAAATCTCACAGGACTT. 2319
1032 elysThrIylGluLeuIylSerThrGluAsnIleSerGlyPheA 1049
2320 .....GCCACCTCAAGCGC 2334
1049 snIylAlaGluIleThrAlaIylAspGlySerAspLeuThrIleGlyAsn 1065
2335 AATCTTAGTGCAGCGGAGACAGCACTATACGTTACGCGCAAGCCG 2384
1066 ThrAsnSerAlaAspIylThrAsnAlaIylSylSValThrPheAsnGluVal 1082
2385 CCAAAAGCGCAACTCAGCCTGTGGGCAATGCCAAGACATTTAATC 2434
1082 lIylAspSerIylSleSerAlaAspIylHisIylSValThrLeuHisSerI 1099
2435 AAGCCACATTTAAAGGCAACATGCGCTTGACACATCTCATTTAAT 2484
1099 ySValIylThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115
2485 CTAAAGCAACGCGCTGACAAAAGCGCAGTGCAGCTTTCGACACGC 2534
1116 AsnAlaGlyLeuThrIleAspAlaIylAsnValThrValAsnAsnI1 1132
2535 TAAGGCAAC.....GTAAGCATTCGCGACACAGCGCANTGTC.... 2574
1132 eThrSerHisIylSAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149
2575 .....TCCCTAGCGGATTAAGCGATATTCATTTTGAACACGC 2613
1149 hIylSThrGlyThrThrIleAsnAlaIleThrThrGlyAsnValGluIleThr 1165
2614 CGCTTACCGGAAATACGCGGCGCAAGATAGC.....GCATT 2654
1166 AlagIleThrIleSerIleLeuGlyIleGluSerSerSerGlySerVal 1182
2655 ACACCTTAAAGACAGCAATGACCGCTGCGGCGGACGAGATTTAGCA 2704
1182 lThrLeuThrAlaThrGluIylAlaLeu.....AlaValSerA 1195
2705 ATTTAAACCTTGACACGCGACCATTAATTCGCTATTCGACAC 2754
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207
2755 GATCGGCGAGCGCGCAACCGGCAAGTGGCGGACAGATCGCGCGCGCG 2804

```

```

1208 .....GlyAlaLeuThrThrLeuAlaGlySer..... 1216
2805 TTGCGCGCTTCCCTATTTATCCGTTACGCCGCACTTCGCGAAGATCC 2854
1217 .ThrIleIylGlyThrGluSerValThr.....ThrSerSerIylSerg 1231
2855 GTTTCAACACGCTGACGTTAAAGCGCAATGTAGCGTACGAGAACAA... 2901
1231 lY.....AspIleGlyGlyThrIleSerGly...GlyThrVal 1242
2902 ...TTCCGCTTATGTGGAACCTTTCGCTACCGCAGCGCAATTTGAA 2948
1243 GluValIylAlaThrGluSerLeuThrThrGluSerIylAsnSerIylLe 1259
2949 GCTGCGGGAAGTTCGAGCGCACTTACACCTTGCTGCTGC..... 2988
1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275
2989 .....AACATACCGCGCAACGACCGCTAAGTCTCGAGCA 3024
1275 lylGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291
3025 TTGACGTA.....GTGAGGAAAGAACAA 3050
1292 LeuThrValAlaIylAsnGlyAlaGluIleAsnAlaThrGluIylAlaIa 1308
3051 CACACCGCTTCGCAAAATCTTAATTCACCTGCAAAAGCAACAGCTCG 3100
1308 lLeuThrThrSerSerGlyIylSerThrThrGluAlaSerSerHisIle 1325
3101 ATGCGGCGGAGTGTATTCAGCTTATTCGCAAGACGCGAGTTCCGC 3150
1325 hrSerAlaIylGlyGluValAsnLeuSerAlaGluAspIylSerValAl 1341
3151 CTGCATATCCGTCAAAGAACAGAGCTTTCGACA 3187
1341 agIylSerIleAsnAlaIaAsnValThrLeuAsnThr 1353

seq_name: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:US-08-469-880-2

seq_documentation block:
; Sequence 2, Application US/08469880
; Patent No. 5876733
; GENERAL INFORMATION:
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: High Molecular Weight Surface Proteins
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Matzare, Ltd.
; STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,880
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9205704.1
; FILING DATE: 16-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/02166
; FILING DATE: 16-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,832

```

FILED DATE: 16-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-516 MIS:Vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-469-880-2

alignment_scores:
Quality: 265.00 Length: 1297
Ratio: 0.438 Gaps: 73
Percent Similarity: 46.646 Percent Identity: 21.126

alignment block:

US-09-303-518D-653 x US-08-469-880-2 ..

Align seg 1/1 to: US-08-469-880-2 from: 1 to: 1536

```
178 AATAAGCAGAGTTTGCAGTCGGCGGCGAAGATATTTAGCTTACACAA 227
||| ||| :||||| ||||| :|||
199 AsnHisIleuIlethrValGly..LysAspLysVal..... 211
228 AAAAGGCGAGTTGGTCGCAATTCGATGACGAAAGCCCGATGATTGAT 277
:|||||:|||||:|||||:
212 .....AsnLeuIleGlyLysValLysAsnGlu..... 221
278 TTCTGTGATGCGCTAACGCG.....GTGGCGCATTTGGCGGCGAT 321
:|||||:|||||:|||||:
222 ..GlyValIleSerValAsnGlyLysLeuLeuAlaGlyGln 237
322 CAA.....TATATGTGAG 335
|||
238 LysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIleAl 254
336 CGTGGCAGATTAAGCGCGCTATAACAATGTTGTTGGTCGGAGGAA 385
:|||||:|||||:|||||:
254 AlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGly 271
386 GCAAT.....CCGATCAGCAGCGCTTTTCT 411
:|||||:|||||:|||||:
271 LysAsnIleAsnValArgAlaIleAlaThrIleArgAsnGlnGlyLysLeuSer 287
412 TACCAATGTGCAAAAGAAATATTAAGCAGGAGCTAACGGCCATCC 461
:|||||:|||||:|||||:
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTAATGCGCGGATTAATCATATGCGCGCTTGCACAAATTTGCACAGAT 511
:|||||:|||||:|||||:
298 .....AsnIleValLeuSerA 303
512 CAGACCTGTTGAG.....ATGACAGTTATATGAT 543
:|||||:|||||:|||||:
303 LAlysGluGlyLysIleGluIleGlyValIleSerAlaGlnAsnGln 319
544 GGGTGAATTAAGCTGATTAATAATACCTGATCGTTT.....CG 587
:|||||:|||||:|||||:
320 GlnAlaLysGlyLysLeuMetIleThrGlyAspLysValThrLeuLys 336
588 AATCGAGCA.....GGCAGACAAATAT 610
:|||||:|||||:|||||:
336 ThrGlyAlaValIleAspLeuSerGlyLysGluGlyLysIleThrTyrL 353
611 GCGGCTGATGAA...GACGAACCAATATC..... 639
```

```
353 euglyIAspGluArgGlyLysAsnGlyLysAsnGlyLysLeuAlaLys 369
:|||||:|||||:|||||:
640 .....CGCGAAAGTTCATATCATATTCAGCAAG..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGly 386
:|||||:|||||:|||||:
667 .....GCATATTCCTTG.....CTGTCGGGCGCATTA 694
||| ||| |||
386 sGlyGlyArgAlaIleValThrPolYAspIleAlaLeuIleAspGlyAsnI 403
695 CTTTGCACAAATTCGATCAGGTGCGCAGCTCAACTTACGTAGCGAA 744
||||| |||||||
403 LeuAlaGln...GlySerGly..... 409
745 AATATTAACATAGCCATATGATGTTTACCAACAGAGGCTCATTTGG 794
:|||||:|||||:|||||:
410 .....AspIleAlaLysThrGlyLysPheValG 419
795 CGACAGTGCCTCACATGTTATCTATGATGCCCAAGCAAGAGTGT 844
||||| |||
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TATTAATGGGATTCGCAACAGGCAACCCCTATATAGAAAAAGCAAT 894
|||||:|||||:|||||:
430 .....AsnAlaIleValAspAla..... 435
895 GCGCTTCACAGCTAGTTCGTAACAGATGTTCTATGATCAAAATCTTGTG 944
|||||:|||||:|||||:
436 .....LysGluThrLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGTATCTACGAAACCATCAAAATGGCAAAATCTTT 994
|||||:|||||:|||||:
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459
995 TTAACAGCAATTAATATATGCGCGAGGAAATTCGATCCAAACATAAAC 1044
:|||||:|||||:|||||:
459 LuAspAspGluThrThrGlySerGly.....AsnSer 469
1045 TATCTCTACCTTATAGATTAATAACAGCAACCGTTCAATGTTAATGT 1094
||| ||| ||| |||
470 AlaSerThrProLysArgAsnLysGluLysThrThr.....LeuThrAsnTh 485
1095 TTCTTTATCCGAG.....ACAGCA 1114
:|||||:|||||:|||||:
485 rThrLeuLysSerIleLeuLysGlyThrPheValAsnIleThrAla 502
1115 GAGAACCTGTTTAT..... 1128
:|||||:|||||:|||||:
502 sngLnaArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518
1129 .....CATCTGCAAGTGGCGTCAACAGTTTCGACC 1160
519 ThrLeuThrPserGluGlyArgSerGlyGlyVal..... 530
1161 CAGACTGAATAT.....GAGAAAAATA 1183
:|||||:|||||:|||||:
531 ..GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547
1184 TTTCCTTTATTCACAAAGAAAGTGAATGATCTATACCAACATC 1233
:|||||:|||||:|||||:
547 eunThrIleTyrSerGlyGlyTyrValAsp.....ValHisLysAsnIle 561
1234 AACCAAGCGCGCGCGGCTTGTATTTGAGGATTAATTTACGGTCGCC 1283
:|||||:|||||:|||||:
562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
1284 TAA.....AACCAAGAACGTCGCAAGCG 1309
|||
573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGly 590
1310 CGGCGCTTCATATCAGTATGATGAGTACCGCTTACTTGAAAGTAAACGCG 1359
|||||:|||||:|||||:
```


590 lnglythr...llethrserglyasnlglylphenarphenasn 605
 1360 GTGGCAACGACGCGCTGTCCAAATGCGCAAGGACGCTGCTGTCA 1409
 606 valser...leuasnglythrlysergly...leuglnpheth 618
 1410 AGCCAAAGGCGGAAAC...CAAGCTCGG 1435
 618 rthrlyearthrlnslystryalalethrlnslyphegluglythr 635
 1436 TCACCGTGGGCGAGGTAAAGTCATCTTA...GAT 1467
 635 euannile...serglyvalasnleisermetvalleuprolysasn 650
 1468 CAGCAGGCGGATCAAGCAAAAACAGCCTTT... 1503
 651 gluserglytyrasplysphenylglyargthrytrpansleuthrse 667
 1504 ...AGTGAATCGGC...TTGCTCAGCGCGCAGG 1531
 667 rleuasnvalsergluserglylupheasnleuthrleaspserrfng 684
 1532 GGACG...GTG 1539
 684 lyseraspserralaglythrleuthrlnprotlyrasnleuasnglyle 700
 1540 CAACGTGAATGCGATATCAAGTTCACCCGACAAA... 1575
 701 serphenasnlyasprthrthrphenasnvalgluarasnalaargvalas 717
 1576 ...CTCT 1579
 717 nphenaspilleysalaproileglyleasnlystryserseurleuasnt 734
 1580 ATTGGGCTTTGCGGCGGACGCTTGTGATTCAGCGCATTCGCTTCG 1629
 734 yralaserphenasnlyasnleaservalserglyglylyseralasp 750
 1630 TTCACCGCGATT...CAAAATACCGATGAAGGCGCATATTGT 1670
 751 phenleuthleualaserseurserasnvalglthrproglyvalal 767
 1671 CAACCAACAT...CAAGCAACAG 1690
 767 easnserlystryrphenasnvalserthrlyserseurleuacpheylyst 784
 1691 AATCCACCGCTTACCATTCACAGC...AATTAAGATATTACT... 1728
 784 hserglyserthrlysthnglypheserleaglulysaspleuthrleu 800
 1729 ...ACAACCGGCAATAC... 1743
 801 asnalathrclyglyasnlethrleuenglvalgluglythrasp 817
 1744 ...AACACTGTGATGCAAAAAGAAATTCGCTACACGCT... 1782
 817 ymetlleglylysglyllevalalalylysasnlethrphlegllyg 834
 1783 ...TGTTTGGCGAGAAAGATGCA...ACCAAAACGAGGCGG 1821
 834 lysanlethrphleglyserarglysalvalthrlnleagllyasn 850
 1822 CTCATGTGAATTCACACCGGAAAGAGCGGATCGCATCTTA... 1863
 851 valthrleasn...asnasnalaasnvalthrleuileglyse 864
 1863 ... 1863
 864 rasphenaspasnhsclnlysproleuthrlelylyasvallel 881
 1864 ...CTGCTTCCGCGGCA...ACAATTTAAACGC 1893
 881 leasnserglyasnleuthrthralaglyglyasnlevalasnillealagly 897

1894 AATATCAGC...CAAAACGCGCAACGTTTTCACGCGCA... 1933
 898 asnleuthrvalgluserasn...phenylsalalethrpnsp 913
 1934 ...GACCGACACCGGACGCTTACATC 1957
 913 hethrphenasnvalglglyleupheaspasnlygllyasnseraslle 929
 1958 ATTTCGAAGCGGGTGTCAAAAATGGAAGATATCCACACAGCAAAATC 2007
 930 serllealalyglylalaargphenylsaspilleaspsnserlyasn 946
 2008 GTGTGGACACAGATGATCGACGC...ACATTTAAAGCGGAAACTTCC 2056
 946 nleuserllethrthrasnserseurthrlyarg... 958
 2057 ATATTCAGGCGGACACAGCGGTGTTCGCCAGATGTCCAAAGTCGAA 2106
 959 ...Thrilleserglyasnlethrlnslyasn 969
 2107 GCGCATTTGCAATTAACATTCACGCC...CAAGCATTTT 2144
 970 glyaspleuasnlethrlnsnglylyserasprthrglumetlnlegl 986
 2145 CGGTGTGCGACCCATCAACAGCCACACATCTGTACAGCTTCGACTGGA 2194
 986 yglyasprvalserclnlysglylulysnleuthrleaserasplyst 1003
 2195 CGGCTGTGACAGTTGTACGAAAAACATTAC...GACGATAAAGTC 2241
 1003 leasnlethr...lyslnlethrlelysalaglyal 1015
 2242 ATTCTTCATTCAGCAACGACGACATGAGGCAATGTACGCTTGCGGA 2291
 1016 aspglyglusnseraspserraspalaThrasnalsnleuthrll 1032
 2292 TCACGCT...CATTTAATCTCACAGACTT... 2319
 1032 elythrlysglylulysnleuthrlnaspleuasnlesercllyphea 1049
 2320 ...GCCACACTCAACGCG 2334
 1049 snlysalaglyllethrlyalaspcllyseraspleuthrthleglyasn 1065
 2335 AATCTTAGTGGAGGCGGACGACACACTATACGCTTACGCGCAACGCGAC 2384
 1066 thrlnsneralaspcllythralsnalaalysvalthrphenasnglnva 1082
 2385 CCAAAACGCGAACCTCAGCTGTGGCAATGCCCAACGCAACATTATATC 2434
 1082 llyaspserrlylleseraleaspglyhlslyvalthrleuhsler 1099
 2435 AAGCCACTTAACGCGCAACACATCGGCTTGGCAATGCTTCAATTAT 2484
 1099 ysvalgluthrserglyserasnasnlnthrlnaspserraspasn 1115
 2486 CTAACCAACACGCGCTACAAACGCGAGCTGACGCTTTCGACACGCG 2534
 1116 asnalaglyleuthrilleaspralalysnvalthrvalasnasnll 1132
 2535 TAAGCAAC...GTAAGCATTTCCGACTACAGCGCAATGTC... 2574
 1132 ethrserhlslysalavalserlleseralahrserglylulieuthr 1149
 2575 ...TCCCTAGCGGATGACGAGTTCATTCATTTGAACAGC 2613
 1149 hrllysthrclythrthrlleasnalaThrthcllyasnvalglulieuthr 1165
 2614 CGCTTACCGGAAAAATCAGCGCGGCAAGATACG...GCAT 2654
 1166 Alaglnthrclyserlleuclglyllegluserseurserglyserva 1182

2655 ACATTAAAGACAGCAATGACGCTGCCGCGGACGGAATATAGCA 2704
1182 lthleuthrlathrglunlglyalaleu.....Alavalsea 1195
2705 ATTAACTTGACACAGCCACCACTTACACTCAATTCGCGCTATGACAC 2754
1195 snlleserGlyasnThrValThrValThrAlaAsnser..... 1207
2755 GATCGGCGGCGGCAACCGGACGAGCGGACGAGCGCGCGCGCG 2804
1208GlyAlaLeuthrThrLeuAlaGlySer..... 1216
2805 TTCCGCGCGCTCCCTATTATTCCTTACGCGCCACCTGCGGAGAAATCC 2854
1217 ThrlelysglyThrGlnserValThr.....ThrserserGlnserG 1231
2855 GTTCAACAGCGTGCAGGTAAAGCGCAATATGACGTCAGGAGACA... 2901
1231 ly.....AspIleGlyGlyThrIleSerGly...GlyThrVal 1242
2902 ...TTCCGCTTATATGTCGAACTTCGCTACCGGACGCGCAATATGA 2948
1243 GlnVallysaIathrglunserLeuthrThrGlnserAsnserlylely 1259
2949 GCTGGCGGAAAGTTCGAAAGCACTTACACTTGGCTGTC..... 2988
1259 s...AlathrThrGlyGlnAlaAsnValThrserAlathrGlyThrIleG 1275
2989AACAAATCCGCGCAACGACCCGCTAGTCTCGAGCA 3024
1275 lyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaIlyasp 1291
3025 TTGACGTA.....GTGGAAGAAAAGACAA 3050
1292 LeuthrValGlyAsnGlyAlaGlnLeuAlaThrGlnGlyAlaAlaIath 1308
3051 CACACCGCTGTCGAAATCTTAATTCACCTGCAGAAAGCAACGCTCG 3100
1308 lreuthrThrserSerGlyLysLeuthrThrGlnAlaSerSerHisIleT 1325
3101 ATGCGCGCGCATGCGCTTATCAGCTTATCCGCAAGACGCGGAGTCCGC 3150
1325 hserAlaIlysglyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341
3151 CTGCATATTCGCTCAAGACAGAGCTTCCGACA 3187
1341 aglySerIleAsnAlaAlaAsnValThrLeuAsnThr 1353

seq_name: /cgn2_6/prodata/1/iaa/5B_COMB.pep:us-08-728-470-2

seq_documentation_block:

Sequence 2, Application US/08728470
Patent No. 5928651

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd.

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/728,470

FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9205704.1
FILING DATE: 16-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Berksstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-633
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
US-08-728-470-2

alignment_scores:
Quality: 265.00 Length: 1297
Ratio: 0.438 Gaps: 73
Percent Similarity: 46.646 Percent Identity: 21.126

alignment_block:
US-09-303-518d-653 x US-08-728-470-2 ..

Align seg 1/1 to: US-08-728-470-2 from: 1 to: 1536

178 AATTAAGCAAGTTGTCAGTCGCGGCGCAAGATATGAGCTTTACAA 227
1199 AsnHisGlyLeuIleThrValGly...LysaspGlySerVal..... 211
228 AAAAGCGAGTTGTCGCGCAATGATGACGAAAGCCCGATGATGAT 277
212AsnLeuIleGlyGlyLysValLysAsnGln..... 221
278 TTCTGTGTAATCGCGTAACGC.....GTGCGGCAATGGCGGCGAT 321
222 ..GlyValIleSerValAsnGlyLysSerIleSerLeuAlaGlyGln 237
322 CAA.....TATATTGAG 335
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIleAl 254
336 CGTGCACATTAACGCGCGCTTAACAATGTTGATTTGTCGAGAGAA 385
254 aAlaProGlnAsnGlnAlaValAsnLeuGlyAspIlePheAlaIlyGly 271
386 GCAAT.....CCGATACGACCCGCTTTCT 411
271 LysAsnIleAsnValAlaAlaIleThrIleArgAsnGlnGlyLysLeuSer 287
412 TACCAATTTGTAAAGAAATTAATTAAGACGAGCACTAAGCGCATCC 461
288 AlaAspSerValSerLysasp.....LysSerGly..... 297
462 TTATGGCGGCGATATATATATGCGCGCTTGCACAATTTGTCACAGATG 511
298AsnIleValLeuSerA 303
512 CAGAACCTGTTGAG.....ATGACGAGTATATATGAT 543
303 lalysGlnGlyGlnAlaGlnIleGlyIlyValIleSerAlaGlnAsnGln 319

544 GGTGGAATACGCTGATTTAAATAATACCGATCGTGT.....CG 587
588 AATCGAGCA.....GGCAGCAATATAT 610
336 sthrglyalavalileaspleuserglylysuglyglylthrlytl 353
611 GCGGCTCGATGAA...GACGACCAATAAC..... 639
333 euglyglyaspeluarnglyluelysasnnglyileglnleualals 369
640CGGAAAGTTCATATTCATATTCAGC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGly 386
667GCATATCTGTG.....CTGCGGTGGCAATA 694
386 sglglylvargalailvalitpplaspilaleuleileaspllyasni 403
695 CTTTGCACAAAATGATCAGGTGGTGACAGTCACACTAGTAGCGAA 744
403 leasnlagin...GlySerGly..... 409
745 AAAATTAACATAGCCCATATGCTTTTACCACAGAGGCTCTTTGG 794
410AspIleAlaLysThrGlyGlyPheValGl 419
795 GCACAGTGGCTCACCATGTTATCTATGATGCCCAAGCAAGTGT 844
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGTATTCAGCAACGCAACCCCTATATAGAAAGCAAT 894
430AsnAlaIleValAspAla..... 435
895 GGCCTTCACCTAGTTCGTAAAGATTGGTCTATGATGAATCTTGTGG 944
436LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGTATCTACGACACATCAAAATGGAAATACCTTT 994
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerg 459
995 TTACGACATTAATAATGGCGCAGGAAATGATGCCAACATTAACAC 1044
459 LAspAspGluTrpThrGlySerGly.....AsnSer 469
1045 TATTCCTACCTATAGATTAATAACAGCAACGCTCAATTTGTTAATG 1094
470 AlSerThrProLysArgAsnLysGluLysThr...LeuThrAsnTh 485
1095 TTCTTATTCGAG.....ACAGCAA 1114
485 rThrLeuGluSerIleLeuLysGlyThrPheValAsnIleThrAla 502
1115 GAGAACTGTTAT..... 1128
502 snGlnArgIleTyValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518
1129CATGCTGACGTGGGTCAACAGTTATGACGC 1160
519 ThrLeuTrpSerGluLysArgSerGlyGlyVal..... 530
1161 CAGACTGAATAT.....CGAGAAATA 1183
531 GlnIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsn 547
1184 TTTCCTTATGACAAAGAAAGTGAATGATCTTACCAACACATC 1233
547 eutThrIleTySerGlyGlyTrpValAsp.....ValHisLysAsnIle 561

1234 AACCAAGCGCGCGGCGGTTTGTATTTTGAAGGCTAAATTTACGCTCGCC 1283
562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
1284 TAAA.....AACACGAAGCTGGCAAGCG 1309
573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590
1310 CCGGCTTCATATACGTGATGGCAGTACCGTTACTTGAAGTAAACGCG 1359
590 lndIlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsn 605
1360 GTGGCAAAACGACCGCTTCACAAATGCGCAAGGCAACGCTGCTGTTA 1409
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618
1410 AGCCAAAGGGAAC.....CAAGGCTCGG 1435
618 rThrLysArgThrAsnLysTyraIleThrAsnLysPheGluGlyThrL 635
1436 TCAGCGTGGCGGACGCTAAGTCATCTTA.....GAT 1467
635 eAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGCGGACGATCAAGCAAAACAAACCTTT..... 1503
651 GluSerGlyTyraAspLysPheLysGlyArgThrTyraAsnLeuThrSe 667
1504AGTGAATCGGC.....TTGTCAGCGGCGAGG 1531
667 rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArg 684
1532 GGAG.....GTG 1539
684 LysSerAspSerAlaGlyThrLeuThrGlnProTyraAsnLeuAsnGlyIle 700
1540 CACTGAATCGCGATATCACTTCAACCCGACAAA..... 1575
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
1576CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTySerSerLeuAsn 734
1580 ATTTCGCTTCGCGCGGACGCTTGGATTGAAACGGGCAATCGCTTTCG 1629
734 yAlaSerPheAsnGlnLysAsnIleSerValSerGlyGlySerValAsp 750
1630 TTCACCGCAT.....CAAAATACGATGAAGGCGGATGATGT 1670
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValIle 767
1671 CAACCAAT.....CAAGCAANAG 1690
767 eAsnSerLysTyraPheAsnValSerThrGlySerSerLeuArgPheLys 784
1691 AATCCACGCTTACATTCAGGC.....AATAAGATAATTC 1728
784 hrSerGlySerThrLysThrGlyPheSerIleGluLysAspLeuThrLeu 800
1729ACACCGGCATATAC..... 1743
801 AsnAlaThrGlyLysAsnIleThrLeuLeuGlnValGluGlyThrAspGl 817
1744AACACTGGATAGCAAAAGAAATTCGCTCAACGCT 1782
817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGly 834
1783TGCTTGGCGGAAGATGCA...ACCAAAACGAACGCGCG 1821
834 LysAsnIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsn 850
1822 CTCATATGTAATTACCAACCGCAAGAAAGCGGATCGCACTTA..... 1863

```

881 ValThrIleAsn.....AsnAsnAlaAsnValThrIleuIleGlySer 864
1863 ..... 1863
864 rAspHeaSpasHisGlnIysProIeuThrIleLysLysAspValIleI 881
1864 .....CTGCTTCCGGCGCA.....ACAATTAAACGGC 1893
881 leuSerSerGlyAsnLeuThrAlaGlyLysnIleValAsnIleAlaGly 897
1894 AATTATCAG...CAACAAACGGCAACTGTTTTCAGCGCA..... 1933
888 AsnLeuThrValGluSerAsn..AlaAsn...PheLysAlaIleThrAsp 913
1934 .....GACGACACCGCAGCCGCTTACAAATC 1957
913 heThrPheAsnValGlyGlyLeuPheAspAsnLysSerLysSerAsnIle 929
1958 ATTAGAGACGGGTGTCATAAAATGAGAGTATCCACAGAGAGAAATC 2007
930 SerIleAlaLysGlyAlaArgPheLysAspIleAspAsnSerLysAs 946
2008 GTGGGACACAGATTGATGACGCGC..ACATTAAACGGCAAACTTCC 2056
946 nLeuSerIleThrThrAsnSerSerThrTyArg..... 958
2057 ATATTCAAGGGCGGACAGCGGTGTTCCCGCAATGTTGCCAAAGTGAA 2106
959 .....ThrIleIleSerGlyAsnIleThrAsnLysAsn 969
2107 GGGGATTGGCATTTAAGCAATCACGCC.....CAAGCATTTT 2144
970 GLyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGlnIleG 986
2145 CGGTGTCGACCGCATCAAGCAACACATCTGTACAGTTCGGGCTGGA 2194
986 yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLys 1003
2195 CGGGTCTGACAGTGTACGAAAAAACCATTACC...GACGATTAAGT 2241
1003 leuAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015
2242 ATGTCTTCATTGACAGACGACGACATCAGAGCATGTACAGCTTGCCGA 2291
1016 AspGlyGluAsnSerAspSerAspAlaThrAsnAlaAsnLeuThrI 1032
2292 TCACGCT.....CAATTAAATCTCACAGACTT. 2319
1032 euThrIleLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPhe 1049
2320 .....GCCACACTGACGGC 2334
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065
2335 AATCTTAGTCAGCGGAGACACGACATATACGTTACGGCGCACGGCAC 2384
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValIleThrPheAsnGlnI 1082
2385 CCAAAACGGCAACTCAGCCTGTGGGCAATGCCAAGCAACATTTATC 2434
1082 lLysAspSerLysIleSerAlaAspGlyHisLysValIleThrHisSer 1099
2435 AAGCACAATTAAAGGCAACACATCGGCTTCGCAATGCTTCATTTAAT 2484
1099 yValGluThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115
2485 CTAAGCAACAAGCCGTACAAAGGACGCTGACGCTTCGACACAGC 2534
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValIleThrValAsnAsn 1132
2535 TAAAGCAAC.....GTAAGCATTCCTCGACACTCAACGGCAATGTC. 2574

```

```

1132 euThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThr 1149
2575 .....TCCTAGCCGATTAAGCAGTATTCATTCATTTGAAAACAGC 2613
1149 hrLysThrGlyThrThrIleAsnAlaThrThrGlnLysValGluIleThr 1165
2614 CGCTTACCGGAAAAATACGGCGCGGCAAGATACG.....GCATT 2654
1166 AlaGlnThrGlySerIleLeuGlyIleGluSerSerSerGlySerAl 1182
2655 ACACCTTAAAGACAGCGAATGACGCTGCCGTGGGACAGGAATTAAGCA 2704
1182 lThrLeuThrAlaThrGluGlyAlaLeu.....AlaValSerAl 1195
2705 ATTTAAACCTTGAACGCCACCATTAACCTACATTCCTGATGACAGC 2754
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207
2755 GATCGCGCAGCGCGCAACCGGACGTGCGCAGATGCGCGCGCGCGCG 2804
1208 .....GlyAlaLeuThrThrIleuAlaGlySer..... 1216
2805 TTGCGCGCTTCCCTATTATCCGTACGCGCGCACTTTCGACAGATCC 2854
1217 .ThrIleLysGlyThrGluSerValThr.....ThrSerSerGlnSer 1231
2855 GTTTCACACACGCTGACGTTAAACGGCAATTGAACGCTCAGGACACA... 2901
1231 Ly.....AspIleGlyGlyThrIleSerGly...GlyThrVal 1242
2902 .....TTCCGCTTATGCGGAGACTCTTCGCTACCGCAGCGCAATTA 2948
1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIle 1259
2949 GCTGCGGAGAAATCGAGGACCTTACACTTGGCTGTC..... 2988
1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIle 1275
2989 .....AACATACCGGCAACGAAACCCGTAAGTCTGAGCA 3024
1275 LyGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGly 1291
3025 TTGACGGTA.....GNGAGAGAAAGACAA 3050
1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAla 1308
3051 CACACCGCTGTCGAAATCTTAATTTCACCGTCAAAAGCAACACGCTG 3100
1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIle 1325
3101 ATCGCGCGCATGGGTTATGAGCTTATCCGCAAAAGACGGCAGATCCG 3150
1325 hrSer..AlaLysGlyGlnValAsnLeuSerAlaGlnAspLysSerVal 1341
3151 CTGCATTAATCCGGTCAAGAAAGCAAGAGCTTCCGACA 3187
1341 agLysIleAsnAlaAlaAsnValThrLeuAsnThr 1353

```

seq_name: /cogn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-617-697-2

seq_documentation_block:

Sequence 2, Application US/08617697

Patent No. 5977336

GENERAL INFORMATION:

APPLICANT: Barenkamp, Stephen J

TITLE OF INVENTION: High Molecular Weight Surface Proteins

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Matlare, Ltd

STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza

SUITE: Bldg. 1

CITY: Arlington

```

STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/617,697
FILING DATE: 01-Apr-1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,832
FILING DATE: 05-Oct-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US93/02166
FILING DATE: 16-Mar-1993
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 1038-557
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 415-0813
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1536 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-617-697-2

```

```

alignment_scores:
Quality: 265.00      Length: 1297
Ratio: 0.438         Gaps: 73
Percent Similarity: 46.646      Percent Identity: 21.126

```

alignment_block:

US-09-303-518D-653 x US-08-617-697-2 ..

```

Align seg 1/1 to: US-08-617-697-2 from: 1 to: 1536

178 AATAAGGCAAGTTTGCACGCGGCGAAGATATGAGCTTACACAA 227
    ||| ||| :|||:||||| ||||| :|||
199 AsnHisGlyLeuIleThrValGly..LysAspGlySerVal..... 211
228 AAAAAAGGAGTTGTGCGCAATCGATGACGAAGCCCGATGATTGATT 277
    :|||:||||| :|||: ||| :|||:
212 .....AsnLeuIleGlyGlyLysValLysAsnGlu..... 221
278 TTTCGTGTATCGCGTACGCGC.....GTGCGGCATTTGCGGCGCAAT 321
    :|||:||||| ||||| :|||: |||||:|||||
222 ..GlyValIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237
322 CAA.....:|||||:|||||:|||||:|||||:|||||:|||||: 335
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
238 LysIleThrIleSerAspIleIleAsnProThrIleThrIleSerIleAl 254
336 CGTGGCAGATACGCGCGCTATACAAATGTTGTTGTCGCGAGGAA 385
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
254 AlaProGluAsnGluAlaValAsnLeuGlyAspIlePheAlaLysGlyG 271
386 GCAAT.....:|||||:|||||:|||||:|||||:|||||:||||| 411
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
271 LysAsnIleAsnValAlaGluAlaThrIleArgAsnGlnGlyLysLeuSer 287
412 TACCAATTTGTAAGAAGAAATATATTAAGACGACGACGACGACATCC 461
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297

```

```

462 TTATGCGCGCATATATATGCGCGTTTGCCAAATTTGTCACAGATG 511
298 .....:|||||:|||||:|||||:|||||:|||||:|||||: 303
512 CAGAACCTGTAG.....ATGACCGATTATATGAT 543
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
303 LysSerGluGlyGluAlaGluIleGlyGlyValIleSerLeuAlaAsnGln 319
544 GGTGGAATATAGCGCTGATTAATAATACCGTCAGTGT.....CG 587
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
320 GluAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLys 336
588 AATCGAGCA.....GCCAGACATATAT 610
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
336 StrGlyAlaValIleAspLeuSerGlyLysGluGlyGluThrTyrL 353
611 GCGGCTGTATGAA...GACGAACCAATATAC..... 639
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
353 euGlyGlyAspGluArgGlyGlyLysAsnGlyIleGluLeuAlaLys 369
640 .....CGCGAAGTTCATATCATATTCAGAAC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGlu 386
667 .....GCATATCTTGG.....CTCGTCGGTGCATTA 694
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
386 sGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGlyAsnI 403
695 CCTTGGCAAAATGATCAGTCAGTGGTGGCACAGTCACTAGTACGAA 744
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
403 LeuAlaAlaGln...GlySerGly..... 409
745 AAATTAACATAGCCCATATGTTTTCACCAAGAGGCTCATTTGG 794
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
410 .....AspIleAlaLysThrGlyGlyPheValGlu 419
795 CGACAGTGCCTACCAATGTTTATCTATGATGCCCAAAACCAAGAGT 844
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTAATGGGTATGTCAAACAGCAACCCCTATATAGAAAAAGCAAT 894
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
430 .....AsnAlaIleValAlaAspAla..... 435
895 GCGTTCAGCTAGTTCGTAAGATGTTGTTCTATGATCAATCTTGCTGG 944
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
436 .....LysGluTrpLeuLeuAsp.....Ph 442
945 AGATACCCATTCAGATATCTACGAACCAACATCAAAATGGAAATACTTT 994
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459
995 TTAACGACAAATATATATGCGCGCAGAGAAATATGATGCCAAACATTAAC 1044
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
459 LuAspAspGluTrpThrGlySerGly.....AsnSer 469
1045 TATTCCTACCTTATAGATTAATAAACAGCAACCGTTCATTTTAATAGT 1094
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
470 AlaSerThrProLysArgAsnLysGluLysThrThr...LeuThrAsnThr 485
1095 TTCTTATATCCGAG.....ACAGCA 1114
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
485 ThrIleuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAla 502
1115 GAGAACCTGTATAT..... 1128
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
502 sGlnIaArgIleThrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518
1129 .....CATGCTGAGGTGGGTCACACAGTTATCGACC 1160
    :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
519 ThrLeuTrpSerGluGlyArgSerGlyGlyVal..... 530
1161 CAGACTGATATAT.....GGAGAAATA 1183

```

```

531 .....|||
    GluLeuAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547
1184 TTTCCTTATTGACAAAGGAAAGTGAATTGATCTTACCAGCAACATC 1233
    .....|||
547 eUrThrIleYrSerIyGlyTrpValAsp.....ValHisLysAsnIle 561
1234 AACCAAGCGCGCGCGTTTGTATTGAGGTATTTTACGTCCTCCGCC 1283
    .....|||
562 SerLeuGlyAla.....GlnLysAsnIleAsnIleThrAl 573
1284 TAA.....AACCAAGCAAGCTGGCAAGCG 1309
    .....|||
573 aLysGlnAspIleAlaPheGlnLysGlySerAsnGlnValIleThrGly 590
1310 CGGCGGTTCATATCATGTATGTCAGTACCGTACTTGAAAGTAAGCGC 1359
    .....|||
590 lngIyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsn 605
1360 GTGGCAAGCGACCGCTGTCGCAAAATGCGCAAGCGACGCTGGTGTCA 1409
    .....|||
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheTh 618
1410 AGCCAAAGGGAAGAAC.....CAAGGCTCGG 1435
    .....|||
618 rThrIlyAsrThrAsnLysTyraIaIleThrAsnLysPheGlnGlyThrL 635
1436 TCACGGTGGGCGGCGTAAAGTCACTTA.....GAT 1467
    .....|||
635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGCGGACGATCAAGGCAAAAAACAAGCCTT..... 1503
651 GluSerGlyTyraSpLysPheLysGlyArgThrTyrrTrpAsnLeuThrSe 667
1504 .....AGTGAATCGGC.....TTGGTCAGCGGCGAGG 1531
667 rLeuAsnValSerGluSerGlyLupPheAsnLeuThrIleAspSerArgG 684
1532 GGACG.....GTG 1539
684 lySerAspSerAlaGlyThrLeuThrGlnProTyraSnLeuAsnGlyIle 700
1540 CAACGTGAATGCCGATATCATGTTCAACCCCGACAAA..... 1575
701 SerPheAsnLysAspThrThrPheAsnValGluArgAsnAlaArgValAs 717
1576 .....CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrrSerSerLeuAsnT 734
1580 ATTTCGGTTTCGCGCGGAGCGATTGATGAGCGGCATTCGCTTCG 1629
    .....|||
734 yAlaSerPheAsnGlnLysAsnIleSerValSerGlyGlySerValAsp 750
1630 TTCACCGCAAT.....CAAAATACGATGAAGGCGGATGATGT 1670
    .....|||
751 PheThrLeuLeuAlaSerSerSerAsnValGlnThrProGlyValValII 767
1671 CAACCAACAT.....CAAGCAAG 1690
    .....|||
767 eAsnSerLysTyrrPheAsnValSerThrGlySerSerLeuArgPheLysT 784
1691 AATTCACCGTTTACCATTTACAGGC.....AATAAGATATTACT... 1728
    .....|||
784 hrSerGlySerThrLysThrGlyPheSerIleGlyLysAspLeuThrLeu 800
1729 ...ACACCGCAATAC..... 1743
801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGlnGlyThrAspGI 817
1744 .....AACAACCTTGATACCAAAAAAGAAATTCCTACACAGCGT. 1782
    .....|||
817 yMetIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGlnGly 834
1783 .....TGCTTGGCGGAAAGATGCA...ACCAAGACAGCGGCGG 1821
    .....|||
834 LysAsnIleThrPheGlySerArgLysAlaValThrGlnIleGlnGlyAsn 850
1822 CTCATCTGAATTATCAACCGGAGAGCGAGATCGACATTTA..... 1863
    .....|||
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864
1863 ..... 1863
864 rAspPheAspAsnHisGlnLysProLeuThrIleLysAspValIleT 881
1864 .....CTGCTTCCGCGCGA...ACAATTAAACGCG 1893
    .....|||
881 LeAsnSerGlyAsnLeuThrAlaGlyLysAsnIleValAsnIleAlaGly 897
1894 AATATCACG...CAACAAAGCGCAAACTGTCTTTCAGCGGCA..... 1933
    .....|||
898 AsnLeuThrValGluSerAsnAlaAsn...PheLysAlaIleThrAsnP 913
1934 .....GACCGACACCGCACGCTTACATC 1957
913 heThrPheAsnValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIle 929
1958 ATTTAGAAGACGGGTGTGTCAAAATGAGAGTATCCCAAGGAGAAATC 2007
    .....|||
930 SerIleAlaLysGlyAlaArgPheLysAspIleAspAsnSerLysAs 946
2008 GTGTGGACACACGATGATGATCGACCGC..ACATTTAAAGCGGAAACTTCC 2056
    .....|||
946 nLeuSerIleThrThrAsnSerSerThrTyraArg..... 958
2057 ATATTCAAGGCGGACAAAGCGGTGTTCGCCGAATGTTCCAAAGTGAA 2106
959 .....ThrIleIleSerGlyAsnIleThrAsnLysAsn 969
2107 GCGGATTTGGCATTTAAGCAATCAGCGC.....CAAGCAGTTT 2144
    .....|||
970 GlyAspLeuAsnIleThrAsnGlnGlySerAspThrGlnMetGlnIleG 986
2145 CGGTTCGACCGGACGATCAAGCCACACAAATGTGTACAGTTGGAGTGA 2194
    .....|||
986 yGlyAspValSerGlnLysGlnGlyAsnLeuThrIleSerSerAspLysI 1003
2195 CGGCTGTACAGATTGTACCGGAAAAACCATACG...GACGATAAAGTG 2241
    .....|||
1003 LeAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015
2242 ATTGCTTCATTGAGCAAGACCGACATCAGAGCAATGTACACCTTGGCGA 2291
    .....|||
1016 AspGlyGlnAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeuThrII 1032
2292 TCACGCT.....CATTTAAATCTCACAGACTT. 2319
    .....|||
1032 eLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPhea 1049
2320 .....GCCACATCAACGCG 2334
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065
2335 AATCTTAGTGAGGCGGAGACACGCACTATACGTTACGGGCAAGCGCAC 2384
    .....|||
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnI 1082
2385 CCAAAAGGCAACGCTCAAGCTGTGGCAATGCCCAAGCAACATTTATAC 2434
    .....|||
1082 LlysAspSerLysIleSerAlaAspGlnLysValThrLeuAsnIleSerI 1099
2435 AAGCCACATTTAAAGCAAGCAACATCGCTTGGGCAATGTTATATTAT 2484
    .....|||
1099 yValGlnThrSerGlySerAsnAsnAsnThrGlnLysPheSerSerAspAsn 1115

```

```

2485 CTAGCAACAACGCGGTACAAAACGCGACGCTTCCGACAAACG 2534
      : : : : : : : : : : : : : : : : : : : : : :
1116 AsnAlaGlyLeuThrIleAspAlaValThrValAsnAsnIle 1132
      : : : : : : : : : : : : : : : : : : : : : :
2535 TAAGGCAAC.....GTAGCCATTCCGACCCACGCAATGTC.... 2574
      : : : : : : : : : : : : : : : : : : : : : :
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluLeuThr 1149
      : : : : : : : : : : : : : : : : : : : : : :
2575 .....TCCGTAGCCGATTAAGCAGATATTCATTTTGAACACG 2613
      : : : : : : : : : : : : : : : : : : : : : :
1149 hLysThrGlyThrThrIleAsnAlaThrThrGlyAsnValGluLeuThr 1165
      : : : : : : : : : : : : : : : : : : : : : :
2614 CGCTTACCGGAAATATCAGCGCGGCGAGATACG.....GCATT 2654
      : : : : : : : : : : : : : : : : : : : : : :
1166 AlaGlnThrGlySerIleLeuGlyGlyLeuSerSerGlySerVal 1182
      : : : : : : : : : : : : : : : : : : : : : :
2655 ACACCTTAAGACAGCAAGATGAGCGCTCCGCGGACGGAATTAAGCA 2704
      : : : : : : : : : : : : : : : : : : : : : :
1182 IThrLeuThrAlaThrThrGluGlyAlaLeu.....AlaValSerA 1195
      : : : : : : : : : : : : : : : : : : : : : :
2705 ATTTAAACCTTGACAAACGCAACCATTAACATTCGCGCTATTCGACAC 2754
      : : : : : : : : : : : : : : : : : : : : : :
1195 snIleSerGlyAsnThrValThrValAlaThrAlaAsnSer..... 1207
      : : : : : : : : : : : : : : : : : : : : : :
2755 GATGGCGGACGCGGCAACCGCGGAGTGGGAGATGGCGCGCGCGCG 2804
      : : : : : : : : : : : : : : : : : : : : : :
1208 .....GlyAlaLeuThrThrLeuAlaGlySer..... 1216
      : : : : : : : : : : : : : : : : : : : : : :
2805 TTCGCGCGGTTCCTTATTCGCTTACGCGCGCAACTTCGCGAGAAATCC 2854
      : : : : : : : : : : : : : : : : : : : : : :
1217 ThrIleLysGlyThrGlnSerValThr.....ThrSerSerGlnSer 1231
      : : : : : : : : : : : : : : : : : : : : : :
2855 GTTTCAACACGCTGACGCGTAACGCGCAATTAAGGTCAGGAGCA... 2901
      : : : : : : : : : : : : : : : : : : : : : :
1231 Ly.....AspIleGlyGlyThrIleSerGly...GlyThrVal 1242
      : : : : : : : : : : : : : : : : : : : : : :
2902 ...TTCGCGTTATGTCGGAACCTTCGCGTACCCGACGCGCAATTCGA 2948
      : : : : : : : : : : : : : : : : : : : : : :
1243 GluValLysAlaThrGlnSerLeuThrThrGlnSerAsnSerLysIle 1259
      : : : : : : : : : : : : : : : : : : : : : :
2949 GCTGCGGGAAGTTCGGAAGGACCTTACCTTGGCTGTC..... 2988
      : : : : : : : : : : : : : : : : : : : : : :
1259 s...AlaThrThrGlyLysValAlaAsnValThrSerAlaThrGlyThrIle 1275
      : : : : : : : : : : : : : : : : : : : : : :
2989 .....AACATACCGGCAACGAACCCGTAGTCTCGAGCA 3024
      : : : : : : : : : : : : : : : : : : : : : :
1275 LysIleThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291
      : : : : : : : : : : : : : : : : : : : : : :
3025 TTGACGCTA.....GTGGAAGGAAMAGACAA 3050
      : : : : : : : : : : : : : : : : : : : : : :
1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaLeuThr 1308
      : : : : : : : : : : : : : : : : : : : : : :
3051 CACACCGCTGTCGAAATCTTAATTCACCTTCGCAAAACGACACGTGC 3100
      : : : : : : : : : : : : : : : : : : : : : :
1308 rLeuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIle 1325
      : : : : : : : : : : : : : : : : : : : : : :
3101 ATGCGCGCGCATGCGCTTATCAGCTTATTCGCAAAAGCGCGAGTTCGCG 3150
      : : : : : : : : : : : : : : : : : : : : : :
1325 hThrAlaLysGlyGlnValAsnLeuSerAlaGlnspGlySerValAl 1341
      : : : : : : : : : : : : : : : : : : : : : :
3151 CTGCAATATCCGTCGCAAGAACAAGACTTTCGACA 3187
      : : : : : : : : : : : : : : : : : : : : : :
1341 aGlySerIleAsnAlaAsnValThrLeuAsnThr 1353

```

seq_name: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:US-08-719-641-2

```

seq_documentation_block:
: Sequence 2, Application us/08719641
: Patent No. 6218141
: GENERAL INFORMATION:
: APPLICANT: Barenkamp, Stephen J
: TITLE OF INVENTION: High Molecular Weight Surface Proteins

```

```

: TITLE OF INVENTION: OF NO. 6218141-Typeable Haemophilus
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESS: Spomaker and Mattare, Ltd.
: STREET: 2001 Jefferson Davis Hwy., 1203 Crystal Plaza
: CITY: Arlington
: STATE: Virginia
: COUNTRY: U.S.A.
: ZIP: 22202-0286
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/719,641
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,832
: FILING DATE: 16-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US PCT/US93/02166
: FILING DATE: 16-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9205704.1
: FILING DATE: 16-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Berkstresser, Jerry W
: REGISTRATION NUMBER: 22,651
: REFERENCE/DOCKET NUMBER: 1038-625
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 415-0810
: TELEFAX: (703) 415-0813
: INFORMATION FOR SEQ. ID NO. 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1536 amino acids
: TYPE: amino acids
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE:
: US-08-719-641-2

alignment_scores:
Quality: 265.00 Length: 1297
Ratio: 0.438 Gaps: 73
Percent Similarity: 46.646 Percent Identity: 21.126

alignment_block:
US-09-303-518D-653 x US-08-719-641-2 ..
Align seg 1/1 to: US-08-719-641-2 from: 1 to: 1536

178 AATAAAGCAAGTTTCAGTCGCGGGAAGATATTGAGTTTCAACAA 227
      ||| ||| : : : : : ||| ||| ||| : : : |||
199 AsnHisGlyLeuIleThrValGly...LysAspGlySerVal..... 211
      : : : : : : : : : : : : : : : : : : : : : :
228 AAAAGGAGATTGTCGCAATTCAGTACGAAACCCCGATGATTGATT 277
      : : : : : : : : : : : : : : : : : : : : : :
212 .....AsnLeuIleGlyGlyLysValLysAsnGlu..... 221
      : : : : : : : : : : : : : : : : : : : : : :
278 TTTCTGTGTTATCGCTTAACGCG.....GTGCGGCAATTGGCGGCGAT 321
      : : : : : : : : : : : : : : : : : : : : : :
222 ..GlyAlaIleSerValAsnGlyGlySerIleSerLeuLeuAlaGlyGln 237
      : : : : : : : : : : : : : : : : : : : : : :
322 CAA.....TATATTGTGAG 335
      : : : : : : : : : : : : : : : : : : : : : :
238 LysIleThrIleSerAspIleLeuAsnProThrIleThrTySerIleAl 254
      : : : : : : : : : : : : : : : : : : : : : :
336 CGTGGCACAATAACGCGGCTATTAACAATGTTGTTGGTCGAGGAGAA 385
      : : : : : : : : : : : : : : : : : : : : : :

```

254 aAlaProGluAsnGluAlaValAsnLeuGlyAspLeuPheAlaLysGly 271
386 GCAATCCGATTCAGACCGCTTCT 411
271 LysAlaIleAsnValArgAlaIleThrIleArgAsnGlnGlyLysLeuSer 287
412 TACCAATGTGAAGAATAATTAATTAAGCAGGACTAAGCCATCC 461
288 AlaAspSerValSerLysAsp.....LysSerGly..... 297
462 TTATGGCGCGATTAATATGCGCGTTGCAACAATTTGTCAAGATG 511
298AsnIleValLeuSerA 303
512 CAGAACCTGTTGAG.....ATGACCACTTAATATGAT 543
303 LAlysGluGlyGluAlaGluIleGlyGlyAlaIleSerAlaGlnAsnGln 319
544 GGGTGAATAATAGCTGATTAATTAATTAACCTGATCGTGT.....CG 587
320 GlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValThrLeuLys 336
588 AATCGGACCA.....GGCAGCAATATT 610
336 sThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluThrTyrL 353
611 GCGCGTCTGATGAA...GACGAACCCATATAC..... 639
353 euGlyLysAspGluArgGlyGlyLysAsnGlyIleGlnLeuAlaLys 369
640CGCGAAGTTCATATCATATGATTCAGAC..... 666
370 LysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysGlyLys 386
667GCATATTCTTGG.....CTCGCGGTGCATA 694
386 sGlyLysArgAlaIleValTyrGlyLysPheAlaLeuIleAspGlyAsnI 403
695 CCTTTCACAAATGATGATCAGGTGCGGACAGTCACTAGTAGCCAA 744
403 LeuAsnIleGln...GlySerGly..... 409
745 AAAATTAAACATAGCCCATATGTTTTTTCACAGAGAGGCTCATTTGG 794
410AspIleAlaLysThrGlyGlyPheValG 419
795 CGACAGTGGCTCACCATGTTTATCTATGATGCCCAAAAGCAAAAGTGT 844
419 uThrSerGlyHisAspLeuPheIleLysAsp..... 429
845 TAATTATGGGTATTGCAACAGCAGCAACCCCTATATAGAAAAAGCAT 894
430AsnAlaIleValAspAla..... 435
895 GGCCTTCACAGTCTGTTGATGATGTTGTTGATGAAATCTTGTGG 944
436LysGluTyrPheLeuAsp.....Phe 442
945 AGATACCATTCATGATTCAGAACACATCAAAATGGAATACGTTT 994
442 eAspAsnValSerIleAsnAlaGluThrAlaGlyArgSerAsnThrSerG 459
995 TTAAGACAAATATATGCGCAGAGAAATATGATGCCAACATCAATACAC 1044
459 LuAspAspGluTyrThrGlySerGly.....AsnSer 469
1045 TATTCCTACCTTATAGATTAAAAACAGAACGCTTCATTTGTTAATGT 1094
470 AlaSerThrProLysArgAsnLysGlyLysThrThr...LeuThrAsnTh 485
1095 TTCTTTATCGAG.....ACAGCAA 1114
485 rThrLeuGluSerIleLeuLysLysGlyThrPheValAsnIleThrAlaA 502
1115 GAGAACCTGTTTAT..... 1128
502 sGlnArgIleTyrValAsnSerSerIleAsnLeuSerAsnGlySerLeu 518
1129CATGTCGAGGTGGGTGCACAGCTATATGACC 1160
519 ThrLeuTyrSerGluGlyArgSerGlyGlyVal..... 530
1161 CAGACTGAAATAT.....GGAGAAATA 1183
531 GluIleAsnAsnAspIleThrThrGlyAspAspThrArgGlyAlaAsnL 547
1184 TTTCCTTATTCACAAAGAAAGGATGATGATTCTTACCAACAATC 1233
547 euThrIleTyrSerGlyGlyTyrValAsp.....ValHisLysAsnIle 561
1234 AACCAAGCGCGCGGCTTTGTTTGAAGGTAATTTACGTCGTCGCC 1283
562 SerLeuGlyAla.....GlnGlyAsnIleAsnIleThrAl 573
1284 TAA.....AACAAAGCAACGTGCAGACGCG 1309
573 aLysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyG 590
1310 CGGCGCTTCATACAGTATGCGAGTACGCTTACTTGGAAATAAACGCG 1359
590 InGlyThr...IleThrSerGlyAsnGlnLysGlyPheArgPheAsnAsn 605
1360 GTGGCAAAAGCAGCCGCTGTCCAAATCGCAAGCAGCAGCTCTGTTCA 1409
606 ValSer.....LeuAsnGlyThrGlySerGly...LeuGlnPheThr 618
1410 AGCCAAAGGGGAAAC.....CAGGCTCGG 1435
618 rThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyThrL 635
1436 TCAGCGTGGGCGACGTAAGTCATCTTA.....GAT 1467
635 euAsnIle...SerGlyLysValAsnIleSerMetValLeuProLysAsn 650
1468 CAGCAGCGGAGATCAAGGCAAAACACAGCTTT..... 1503
651 GluSerGlyTyrAspLysPheLysGlyValArgThrTyrTrpAsnLeuThrSe 667
1504AGTGAATCGGC.....TTGTCAGCGCGCAGCG 1531
667 rLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerArg 684
1532 GGACG.....GTG 1539
684 LysSerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyIle 700
1540 CAACGATGCGCGATATACATGATCAACCCGACAA..... 1575
701 SerPheAsnLysAspPThrPheAsnValGluValArgAsnAlaArgValAs 717
1576CTCT 1579
717 nPheAspIleLysAlaProIleGlyIleAsnLysTyrSerSerLeuAsnT 734
1580 ATTTCGCTTTCCGCGCGAGCGATTGATTTGAACGGGCAATCGCTTTCG 1629
734 rAlaSerPheAsnGlyAsnIleSerValSerGlyGlyLysSerValAsp 750
1630 TTCACCGCAT.....CAAAATACCGATGAAGGCGCATGTGT 1670
751 PheThrLeuLeuAlaSerSerSerAsnValGluThrProGlyValAlaI 767
1671 CAACCAACAT.....CAGACAAAG 1690
767 eAsnSerLysTyrPheAsnValSerThrGlySerSerLeuArgPheLysT 784


```
1691 AATCCACCGTTACCATTTACAGC.....ATAAAGATATTACT... 1728
1729 ...ACAACCGCGAATTAAC..... 1743
801 AsnAlaThrGlyGlyAsnIleThrLeuLeuGlnValGluGlyThrAspG1 817
1744 .....ACCACTTGATAGCAAAAAAGAAATTCCTACACGGT. 1782
817 yMeIleGlyLysGlyIleValAlaLysLysAsnIleThrPheGluGlyG 834
1783 .....TGCTTGGCGGAGCAAGATCA...ACCAAAACGACGGGCGG 1821
834 LysAsnIleThrPheGlySerArgLysAlaValThrGluIleGluGlyAsn 850
1822 CTCATCTGAAATTACCAACGAGAGCGGATCGCACTTTA..... 1863
851 ValThrIleAsn.....AsnAsnAlaAsnValThrLeuIleGlySe 864
1863 ..... 1863
864 RasPheAspAsnHisGlnLysProLeuThrIleLysLysAspValIleI 881
1864 .....CTGCTTCCGCGCA.....ACAAATTTAAACGGC 1893
881 LeAsnSerGlyAsnLeuThrAlaGlyLysAsnIleValAsnIleAlaGly 897
1894 AATATCAG...CAACAACGCGCAAACTGTTTTCACGCGCA..... 1933
898 AsnLeuThrValGluSerAsnAlaAsn...PheLysAlaIleThrAsn 913
1934 .....GACGACACCGCGACGCTTCAATC 1957
913 heThrPheAsnValGlyLysLeuPheAspAsnLysGlyAsnSerAsnIle 929
1938 ATTTCGAGACGGGTGTCAAAATGGAAGATCCCAAGGAGAAATC 2007
930 SerIleAlaLysGlyAlaArgPheLysAspIleAspAsnSerLysAs 946
2008 GTGTGGGACACGATTCGATCGACGC..ACATTTAAAGCGGAAAACTTC 2056
946 IleuSerIleThrThrAsnSerSerSerThrIArg..... 958
2057 AATATTCAGGCGGACACGCGTGTTCGCCAATGTTGCCAAATGGA 2106
959 .....ThrIleIleSerGlyAsnIleThrAsnLysAsn 969
2107 GCGGATTGCAATTAAAGCAATCAGCC.....CAAGCAGTTT 2144
970 GlyAspLeuAsnIleThrAsnGluGlySerAspThrGluMetGlnIleG 986
2145 CGGTGTCGACCGCATCAAGCACACAACTGTACACGTTGCGACTGA 2194
986 yGlyAspValSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysI 1003
2195 CGGGTCTGACAAGTTTACCGAAAAACATTACC..GACCATTAAGG 2241
1003 LeAsnIleThr.....LysGlnIleThrIleLysAlaGlyVal 1015
2242 ATTCGCTCATTCAGCAAGACGACATCAGAGCAATGTCAGCTTCGCA 2291
1016 AspGlyLysAsnSerAspSerAspAlaThrAsnAlaAsnAlaAsnLeuThrI 1032
2292 TCACGCT.....CATTTAAATTCACGAGCTT. 2319
1032 elyThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPheA 1049
2320 .....GCCACATTCACAGCG 2334
1049 snLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyAsn 1065
2335 AATCTTAGTCAGCGGAGACAGCACTATACGGTTACGGCAACGCGCAC 2384

::: ||||| ||| ::: ||||| ||| :::
1066 ThrAsnSerAlaAspGlyThrAsnAlaLysLysValThrPheAsnGlnVal 1082
2385 CCAAAACGGCAACCTCAGCCTCTGCGCAATGCCAATGCCAATTTATTC 2434
1082 LlysAspSerLysIleSerAlaAspGlyHisLysValThrLeuHisSerL 1099
2435 AAGCCACATTAACGCAACACATCGCTTCGACATGCTTCATTTAAT 2484
1099 yValGluThrSerGlySerAsnAsnAsnThrGluAspSerSerAspAsn 1115
2485 CTAAACACACACCGCTACAAAACGCGAGTCTGACGCTTTCCGACAAAGC 2534
1116 AsnAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAl 1132
2535 TAAAGCAAC.....GTAAAGCATTCGCGCTCAACGCGCATGTC... 2574
1132 eThrSerHisLysAlaValSerIleSerAlaThrSerGlyGluIleThrT 1149
2575 .....TCCCTAGCCGATTAAGCAGATATTCATTTGAAAAACGC 2613
1149 hLysThrGlyThrThrIleAsnAlaIleThrGlyLysAsnValGluIleThr 1165
2614 CGCTTACCGGAAAAATCAGCGCGCGCAAGATACG.....GCATT 2654
1166 AlaGlnThrGlySerIleLeuGlyGlyIleGluSerSerSerGlySerVal 1182
2655 ACACCTTAAAGACAGGAATGAGCGCTGCCGTCGCGGACGAGAAATAGCA 2704
1182 IThrIleThrAlaThrGluIleValLeu.....AlaValSerA 1195
2705 ATTTAAACCTTGACAAAGCCACCATTAACCTCAATTCGCTTACGACAC 2754
1195 snIleSerGlyAsnThrValThrValThrAlaAsnSer..... 1207
2755 GATCGCGGACGCGCGCAACCGCAGTGGCGGAGATCGCGCGCGCGC 2804
1208 .....GlyAlaLeuThrThrLeuAlaGlySer..... 1216
2805 TTCGCGCGGTTCCCTATTATCCGTTACCGCGCAACTTCGCGCAATCC 2854
1217 ThrIleLysGlyThrGluSerValThr.....ThrSerSerGlnSerG 1231
2855 GTTTCACACGCTGACGTAACGCGCAAAATGAAGCTCAGGACAA... 2901
1231 Ly.....AspIleGlyIleThrIleSerGly...GlyThrVal 1242
2902 ...TTCGCTTATGTCGGAACCTCTGCGCTACCGGACGCGCAATGGA 2948
1243 GluValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIle 1259
2949 GCTGCGGAAAGTTCCGAAAGCACTTACACCTTGCTGTC..... 2988
1259 s...AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleG 1275
2989 .....AACATACCGGCAACGAAACCCGTAAGCTCGAGCAA 3024
1275 LysGlyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAsp 1291
3025 TTGACGGGA.....GTGAGAGGAAAAAGCAAA 3050
1292 LeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThr 1308
1308 IleuThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleT 1325
3051 CACACCGCTGTCGAAATCTTAATTCACCTGCAAAAACACACGTCG 3100
3101 ATCCGCGCGCATGCGGTATACGTTATCCGGAAGACGCGAGTCCG 3150
1325 hrSerAlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAl 1341
3151 CTGCATATCCGTCAAAGAACAGAGCTTTCGACA 3187
```

```

1341 aGlySerIleAsnAlaIAsnValThrIleuAsnThr 1353
seq_name: /cgn2_6/prodata/1/1aa/6B_COMB.pap:US-09-268-347-36
seq_documentation block:
; Sequence 36, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentl In Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2411
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-268-347-36

alignment_scores:
Quality: 257.50      Length: 1701
Ratio: 0.341         Gaps: 75
Percent Similarity: 44.444      Percent Identity: 18.871

alignment_block:
US-09-303-518D-653 x US-09-268-347-36 ..
Align seg 1/1 to: US-09-268-347-36 from: 1 to: 2411

7 ACAACGCAAAAGCGACACCA.....ACACCGCAAGCCCC 47
||| |||||:||||: ||| :||| :|||:
242 ThnGlyAspLysAsnThrIleuAspValIleuThrIleuValGluAsnGlu 258
||| |||||:||||: ||| :||| :|||:
48 TAAACGCGCGCCATCCGCTTCGCCCGCTTACTTACCATATGCGCCGT 97
||||| :|||:||||: ||| :||| :|||:
258 yLysThrIleuValIleuValIleuValIleuValIleuValIleuVal 275
||||| :|||:||||: ||| :||| :|||:
98 CGTGGCGCTTCTGCCCAAGCGCGGCGGACACATTTATTTGGCGATC 147
||| ||| ||| :|||: ||| :||| :|||:
275 yAspIleuValIleuValIleuValIleuValIleuValIleuVal 291
||| ||| ||| :|||: ||| :||| :|||:
148 AACCTACCAATCTATCGGACTTGGCGCGCAAAATAAGCAAGTTGCAGT 197
||| ||| :|||: ||| :||| :|||:
292 ThrSerAsnThrIleuValIleuValIleuValIleuValIleuVal 308
||| ||| :|||: ||| :||| :|||:
198 CGGCGCGCAAAAGATATTTAGGTTTACACCAAAAGGAGGAGTTGGCGCA 247
||| ||| :|||: ||| :||| :|||:
308 rAlaIleuValIleuValIleuValIleuValIleuValIleuVal 325
||| ||| :|||: ||| :||| :|||:
248 AATGCATACCAAAAGCGCGCGATGATGATTTTCTGTGTATGCGCGT 294
||| ||| :|||: ||| :||| :|||:
325 hTrhIleuValIleuValIleuValIleuValIleuValIleuVal 341
||| ||| :|||: ||| :||| :|||:
295 AACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 344
||| ||| :|||: ||| :||| :|||:
342 ThrAsnValIleuValIleuValIleuValIleuValIleuValIleuVal 358
||| ||| :|||: ||| :||| :|||:
345 TACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 394
||| ||| :|||: ||| :||| :|||:
358 sAspThrAsnIleuValIleuValIleuValIleuValIleuValIleuVal 364
||| ||| :|||: ||| :||| :|||:
395 ATCAGCAACCGCTTTTACCAATTTGAAAGAAATATTAATAAGCA 444
||| ||| :|||: ||| :||| :|||:
365 .....IleThrValIleuValIleuValIleuValIleuValIleuVal 373
||| ||| :|||: ||| :||| :|||:
445 GCGCATACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 494
||| ||| :|||: ||| :||| :|||:
374 Gly...AspGlyLeuValIleuValIleuValIleuValIleuValIleuVal 383
||| ||| :|||: ||| :||| :|||:
495 CAAATTTGTCAGATGCAAGACCGCTTGTGATGACCGATTAATGATGATG 544
||| ||| :|||: ||| :||| :|||:

```

```

635 rSerLysSerGlu.....ASnGLyShSt 644
1274 CGGTCGCGCTAAACAAACAGAAACGTGGCAAGCGCGCGCTTCATATC 1323
      |||||
644 hrLleThrValSerValAlaGluThrLysAlaAspSerGlyLeuGlu... 659
      |||||
1324 AGCGATGGCAGTACCGCTTACTTGGAAAGTAAAGCGCGGCAACGACGC 1373
      |||||
660 LysAspLysAspThrLLeLysLeuLysValAspSngInsnThrAspAs 676
      |||||
1374 CCGTCGCAAAATCGGC...AAAGCAGCGTCGCTGGTTCAGCCAAAGGGG 1420
      |||||
676 nValLeuThrValGLeLysnAsnGLyThrAlaValThrLysGLyPheG 693
      |||||
1421 AA.....AMCCAAAGCTCGGTCAACCGTGGCGACGATAAATC...ATC 1461
      |||||
693 LuthrValLysThyGLeYAlaThrAspAlaAspArgLysValThrVal 709
      |||||
1462 TTATATACGAGCGGAGCATCAAGGCAAAACAAACAGCC...TTTATGTGA 1508
      |||||
710 LysAspAlaThrAlaAsnAspAlaAspLysValAlaThrValLysAs 726
      |||||
1509 AATCGCGCTGGTCAGCGGCGAGCGGACGCTGCACACTGAATGCCGATAT 1557
      |||||
726 pValAlaThrAlaLeuAsnSerAlaAlaThrPheValLysThrGluAsnL 743
      |||||
1558 .....CAGTTCACAGCC.....GAC 1572
      |||||
743 eutThrSerLeAspLunAspAsnProThAspSngLysAspAsp 759
      |||||
1573 AAACCTATATTCGCGC.....TTTCGCGGCGGA.....CG 1601
      |||||
760 AlaLeuLysAlaGluAspThrLeuThrPheLysAlaGluLysAsnLeuL 776
      |||||
1602 TTTCGATTTGAACGGCGCATTCGCTTTCGTTCCACCGCATTCAAATACCG 1651
      |||||
776 sValLysArgAspGlyLysAsnLLeThrPheAspLeuAlaLysAsnLeuG 793
      |||||
1652 ATGAGGGGCGATGATTCGCAACCAACATCAAGACAAAGATCCACCGTT 1701
      |||||
793 LuValLysThrAlaLysValSer.....AspThrLeu 803
      |||||
1702 ACCATTACAGCAATAA..... 1719
      |||||
804 ThrLLeGlyGLyLysnThrProThrGlyGlyThrThraLathrProLysVa 820
      |||||
1720 GATATTACTACACCGCAATATACAAACAATTG..... 1752
      |||||
820 LAsnLLeThrSerThrAlaAspLysLeuAsnPheAlaLysGluThraLaa 837
      |||||
1753 .....GATAGCAAAAAGAAATTCCTACACGCT.....TGCTTT 1788
      |||||
837 sPalAserLysSerLysAsnValThrLeuLysGlyLeaLathrThrLeu 853
      |||||
1789 GCGGACAAAGATGACCAACAAACAGACGGCGGCTCATCTGATTAACA 1838
      |||||
854 ThrGluProSerAlaGluLysSerSerHisValAspLeuAsnValAs 870
      |||||
1839 ACCGAGAAGACGGATCGACCTTA.....CTGCTTTCGGCGC 1876
      |||||
870 palatnLrLysSerSAsnAlaLaserLLeGluAspValLeuArgLag 887
      |||||
1877 GAACAAATTTAAACGCAAT..... 1896
      |||||
1897 .....AT 1896
      |||||
904 AspThrValAsnPheThrAspSerThrGlyThrThrThraLathrVa 920
      |||||
1899 CACGCAACA..... 1906
      |||||
920 lThrGluLysAlaAspLysGlyValaAspValLysLLeGluLalalysT 937

```

```

1909 ..... AACGGCAAACTGTTTTCACGGCGCAACGG 1938
1937 hrgelValIlelyAsprhlnslnlyLysleu...PhehrgIlyLysAsp 952
1939 AACACGGACGCGCAACATCAATTGAGAAAGCGGGTGGCCAAATAATGAAAG 1988
953 leuLysAsprhlnslnsn...GlyAlaIthValSerGluAspArg 967
1989 TATCCSCAAGAGAA...ATCGTGTGGCAACGATGGATGCAACGCCA 2035
967 yLysAspThrgIlyhrgIyLeuValIthrlAlaLysThValIleAspAlav 984
2036 CATTTAAAGCGGAAACTTCATATTCAGCGCGGCAACAGCGGTGTTTCC 2085
984 alsnLysSer...GlyTrrIyValIthrgIyGluGlyAlaIthValleu 999
2086 CGCAATGTGGCCAAAGTGGAAAGCGGATGGCATTTAAGCAATCAACGCCCA 2135
1000 ThrlAlaIthrlAlaVal.....AsnIaGlyAsnIaGluTh 1012
2136 AGCAGTTTTCGGTGTGCA.....CGCATCAAGCCACACAACT 2176
1012 rValIthSerGlyIthSerValAsnPhelyAsnIyAsnIaIthThra 1029
2177 GTCAACGTTGGCATGGACGGGT..... 2199
1029 IatrhValSerLysAspAsnIyAsnIleAsnValLysTyAspValasn 1045
2200 .....CTGCACAAGTTG 2210
1046 ValGlyAspGlyLeuLysIleGlyAspAspLysLysIleValAlaIasprh 1062
2211 TACGCAAAAACCATTACCGAGCAAGTAAGTATGTCCTCATTCAGCAAGA 2260
1062 TrhTrhLeuThraIthrgIyLysValSerValProIaIaGlyAla 1079
2261 CGGACATCAGAGGCAATGTCAGCGCTGGCATCAGCGCTATTAACTTC 2310
1079 snSerValAsnAsnIyLysleu.....ValAsnAla 1090
2311 ACAGACTTGGCCACACTCAAGCGCAACTTAGT..... 2343
1091 GluGlyLeuAlaIthrlAlaLeuAsnIleuSerTrpThrlAlaLysAlas 1107
2344 .....GCAAGCGGAGACACGCATATACGGTT... 2370
1107 pLysTyrlAlaAspGlyLysSerGluIyGlyIthAspGlnGluValLys 1124
2371 .....ACGGCGCAACGCCACCAAAACGGCAACTCAGCGTCGCGGCAAT 2415
1124 IaGlyAspLysValIthrlPhelyAlaGlyLysAsnLeuLysValLysGln 1140
2416 GCCCAAGCAACATTAT.....CAAGCACATTAAACGGCAAC 2456
1141 SerGluLysAspPhehTrhYrSerLeuGlnAspThrlLeuThrgIyLeuTh 1157
2457 ATCGCGTTCGGACAATGCTTCATTATCTAACCAACAAC.....G 2497
1157 rSerIleThrlLeuGlyGlyThrlAlaAsnIyLyrAsnAspThrgIyThrv 1174
2498 CGGTACAAAACGGCAGTCAAGCTTTCGACAAACGCTAAGGCAACGTA 2547
1174 alIleAsnLysAspGlyLeuThIleThrlLeuAlaAsnGlyAlaIaIa 1190
2548 AGGCATTTCGCACTCAAGCGCAATGTCTCCCTAGCCATAAGCAGATAT 2597
1191 GlyThraspLaserAnGlyAsnThr..... 1199
2598 CCAATTTGAAACAGCGCGTTTACCGCAAAAATTCAGCGCGCAAGATA 2647
1200 .....IleSerValThrlLysAsp... 1205

```

```

2648 CGGCACTTAAAGACGGAATGACGCTCCGCGGACGGA 2697
1206 .....GlyIleSer 1208
2698 TTAGCAATTTAACTTGACAAACGACATTAACCTATTCGCGCTA 2747
1209 AlaGlysnLysGluIleThrAsnValLysSerAlaLeuLysThr...Ty 1224
2748 TCGA.....CAGCATGCGGACGCGGCAAAACCGGACGTGCGCGAGAG 2791
1224 rLysAspThrGlnAsnThrAlaGlyAlaThrGlnProAlaAlaAsnThrA 1241
2792 CGCGGCGCGCGCGCTGCGCGCGCTTCCCTATTATTCGTCGCGCACT 2841
1241 La.....GluValAlaLysGlnAspLeuValAspLeuThrLysProAla 1255
2842 TCG.....GCAGATCCCGCTTCAACACGCT 2867
1256 ThrGlyAlaAlaGlysnGlyAlaAspAlaLysAlaProAspThrThrAl 1272
2868 GACGCTAAACGCGCAATTAACGCGTCAAGCAATTCGCTTTATGTCG 2917
1272 aaLathValGlyAspLeuArgGlyLeuGly..... 1282
2918 AACTCTCGGCTACCGGACGCAATTAAGCTGCGGAAAGTTCGGA 2967
1283 .....TrpValLeuSerAlaLysLysThrAlaAspGluThrGlnAsn 1296
2968 GGCACCTTACACCTTGCTGTCAACATTAACGCGCAAC...GAACCGGTAG 3014
1297 LysGluPheHisAlaAlaValLysAsnAlaAsnGluValGluPheValGlu 1313
3015 TCTCGAGCAATTGACGCTAGTGGAAAGAAAGAACACACACCGCTGTCCG 3064
1313 yLysAsnGlyAlaThrValSerAlaLysThrAspAsn..... 1325
3065 AAATCTTAATTTCAACCTGCAAAACGAACGCTGATCCGCGCGCATAG 3114
1325 ..... 1325
3115 CGTTATACGCTTATCCGCAACGCGAGTTCGCTCATTAATCCGCT 3164
1326 .....AsnGlyLysHisThrValThrIleAspVal 1335
3165 CAAGACAAAGAGCTTCCGCAAACTGCGCAAGCGGAGAGAAACAGAG 3214
1335 LaGluAlaLysValGlyAspGlyLeuGluLys.....AspThrAspG 1350
3215 CGCCTTGACGCGCAAA..... 3231
1350 yLysIleLeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThr 1366
3232 ...CAGGCAACAATTGCGCGCAAAACAGCGGAAAGAACAGCGCA 3278
1367 ValAspAlaThrLysGlyAlaSerValAlaLysGlyLupheAsnAlaVala 1383
3279 AAGCTTACGCGCTGATTCGCGCGCGGCGCAATGCGCGCAAGAGAGAG 3328
1383 lThrThrAspAlaThrThrAlaGlnGlyThrAsnAlaAsnGluArgGlyL 1400
3329 AAAGTGTTCGGAACGCGCGCGGACGAGCGAATAATCCGCGCATT 3378
1400 yValValValLys.....GlySerAsn...GlyAla 1409
3379 ATGACGCGGCGGAAGAAAGAAAGAGGTGCGCGGATTAAGACAGCGC 3428
1410 ThrAlaThrGluThrAspLysLysValAlaLThrValGlyAspValAla 1426
3429 CTTGCGCAAAACAGCGCGAAGCG.....GAACCGCGCGCGC 3463
1426 alYsAlaIleAsnAspAlaAlaThrPheValLysValGluAsnAspAspS 1443
3464 CTACACCGCGCTTCCCGCGCGCGCGCGCGCGGATTTGCCGCA 3513
1443 ..... 1451
3514 CGGAGCGCCCAACCGCAACCCGCAACCGGACGCGCATCATACCGCTTA 3563
1452 AspAspLysAlaAsnAspAlaLeuLysAlaGlyAspThrLeuThrLeuLys 1468
3564 TGCCATATAGCGGTTTGAAGTAATTTCCGCAACGCTCAACACAGCTTTCG 3613
1468 saLagLysAsnLeuLysValLysArgAspGlyLysAsnIleThrPheA 1485
3614 CCGTACAGAGCAATTAAGACCGCGGTGTTCCGGAAGACCGCGCAAGCC 3663
1485 laLeuAlaAsnAspLeu.....Ser 1491
3664 GTTGGACAGCGCATCCGCGACACCAACACTACCGTTGCGAAGATT 3713
1492 ValLysSerAlaThrValSerAsp..... 1499
3714 CCGCGCTTACCGCAACACCGACCTGCGCAATTCGATGCAAA 3763
1500 .....LysLeuSerLeuGlyThrAsnGlysnLysVala 1511
3764 ACCTCGCAGCGCGCGCTGCGCATCTGTTTCG...CACACCGGACC 3810
1511 snIleThrSerAspThrLysGlyLeuAsnPheAlaLysAspSerLysThr 1527
3811 GCA.....AACACTTCGACGAGCGCATCGGCAACTCGGACGCGCT 3851
1528 GlyAspAspAlaAsnIleHisLeuAsnGlyIleAlaSerThrLeuThr.. 1543
3852 TGCCACGCGTGCCTTTTCGGGCATATACGCATGCGAGTTCACATCG 3901
1544 .....AspThrL 1546
3902 GCATACGCGCGCGCGGTTTGTAGTACGCGCAGCTTTCAGACGATC 3951
1546 euLeuAsnSerGlyAlaThrThrAsnLeuGly.....GlyAsnGlyIle 1560
3952 AAGGCAAAATCCGCGCGCGCTGCTGATTAACGATTCAGGCAAGATA 4001
1561 ThrAspAsnGluLysArgAlaAla.....SerValLysAspValle 1575
4002 CGCGCAGGTTTCGGCGGATTCGGCATCGAACCACATCGCGGCAAGC 4051
1575 uAsnAlaGlyThrAsnValArgGlyValLysPro..... 1586
4052 GCTATTTGTCCAAAAGCGGATTACGATACGAAACGTCATATTCGCC 4101
1587 .....AlaSerAlaAsnAsnGlnValGluAsnIleAspPheVal 1599
4102 ACCCGGCGCTTGATTCACCGCTACCGCGCGGCAATTAAAGCAGATTA 4151
1600 Ala..... 1600
4152 TTCAATTAACCGCGCAACACATTCATCATCGCTTATTGAGCCTGT 4201
1601 .....ThrTyra 1603
4202 CCTTACCGATGCGCTTCGCGCAAAAGTCCGAAGCGGCTCATACCGC 4251
1603 spThrValAspPheValSerGlyAspLysAspThr...ThrSerValThr 1618
4252 GTATTGCGGAGATTTTCGCAAAACCGGACGCGGATGGCGGCTTAA 4301
1619 ValGluSerLysAspAsnGlyLysArgThrGluValLysIleGlyAlaLys 1635
4302 CGCGGAATC.....AAAGTTTCAAGCTGTCCCTCC 4333
1635 sThrSerValIleLysAspHisAsnGlyLysLeuPheThr..... 1648
4334 ACCTGCGCGCGCAAGGCGCAATTTGAAAGCGGACGCGGCGC 4383

```

1649GlyLysGluLeuLysAspAlaAsnAsnGly 1659
 4384 ATC 4386
 1660 Val 1660

seq_name: /cgn2_6/plodata/1/1aa/6B.COMB.pep:US-09-268-347-47

seq_documentation block:

; Sequence 47, Application US/09268347
 ; Patent No. 6335182
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
 ; FILE REFERENCE: 1038-860
 ; CURRENT APPLICATION NUMBER: US/09/268,347
 ; CURRENT FILING DATE: 1999-03-16
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: Patentl Ver. 2.0
 ; SEQ ID NO 47
 ; LENGTH: 2354
 ; TYPE: PRP
 ; ORGANISM: Haemophilus influenzae
 ; US-09-268-347-47

alignment_scores:
 Quality: 254.50 Length: 1664
 Ratio: 0.346 Gaps: 75
 Percent Similarity: 44.231 Percent Identity: 19.591

alignment_block:
 US-09-303-518d-653 x US-09-268-347-47 ..

Align seg 1/1 to: US-09-268-347-47 from: 1 to: 2354

```

7 ACACCGCAACGACGACGCA.....ACACCGCAACGACCGCC 47
  ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
246 ThtGlyAspLysAsnThrLeuAspAlaValLeuThrAlaLysGluAsnGly 262
  ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
48 TAAACCGCGCGCATCGCTCTCGCCCTTACTTACTTACCATATGCTGT 97
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
262 ylsthrThrGluValLysPheThrProLysThrSerValLleLysGluL 279
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
98 CGTTCGCGATTCGCCCCACGCCGCGCGGACACACTTATTTCCGCAATC 147
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
279 ysAspGlyLysLeuPheThrGlyLysGluAsnAsnAspThrAsnLysVal 295
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
148 AACTACCAATACTATTCGCGACTTTCGCAAAATAAAGCAAGTTTCAGT 197
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
296 ThsSerAsnThrAlaThrAspAsnThrAspGlyLysGlyLysValTh 312
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
198 CGGGGCGAAGATTTGAGGTTTACAAACAAAAAGGAGTTCGTGCGCA 247
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
312 rAlaLysAlaValLleAspAlaValAsnLysAlaGlyTTParGValLysT 329
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 AATCGATGACGAAGCCCGCATGATGATTT..TCTGTGATATGCGGT 294
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
329 hrThrThrAlaAsnGlyLysAsnGlyAspPheAlaThrValAlaSerGly 345
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
295 AAGCGCGTGGCGGCGATTCGCGGCGATCATATATTGTGAGCGTGCACA 344
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
346 ThrAsnValThrPheGluSerGlyAspGlyThrThrAlaSerValThrL 362
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
345 TAAACGCGCGCTATACAAATGTTGATTTGTCGCGAGGAGCAATCCCG 394
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
362 sAspThrAsnGlyAsnGly..... 368
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
395 ATCGACACCGCGTTTCTTACCAATTTGAAAAGAAATATTAATAAGCA 444
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
369 .....LleThrValLysTyrAspAlaLysVal 377
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 GGGACTAAGCGCATCCTTATGGCGGCGATATCATATATGCCGCTTGCA 494
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
378 Gly...AspGlyLeuLysPheAspSerAsp.....Lys 387
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
495 CAATTTTGTACAGATGCAAGACCTGTGTAGATGACCGACTATATGATG 544
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
387 sLysLleValAlaAspThrThrAlaLeuThrValThr.....G 400
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545 GGTGGAATACGCTGATTTAATTAATACCGTCGATGCTGTGCAATC... 591
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
400 LysLysValAlaGluLleAlaLysGluAspAspLysLysLysLeuVal 416
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
592 .....GGAGCGCGAGCAACATATTGCGG..... 615
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
417 AsnAlaGlyAspLeuValThrAlaLeuGlyAsnLeuSerTrpLysAlaL 433
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
616 .....TCTGATGAGAC.....GACCCCAATAACCGCGAAGTT 649
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
433 sAlaGluAlaAspThrAspGlyAlaLeuGluGlyLysSerLysAspGlu 450
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
650 CATATCATATTGCAAGCGCATATTCTGGCTGCGCGGCAAT...ACC 696
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
450 LuValLysAlaGlyGluThrValThrPheLysAlaGlyLysAsnLeuL 466
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
697 TTGCACAAAATGATCAGT..... 717
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
467 ValLysGluAspGlyAlaAsnPheThrThrSerLeuGluAspAlaLeuTh 483
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
718 .....GTGCGCACGTCACACTTATGAGGCAAA 745
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
483 rGlyLeuThrSerLleThrLeuGlyGlyThrThrAsnGlyLysAsnAsp 500
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
746 AA.....ATTAACATAGCCCATATGTTTTCACCAAGAGAGGC 786
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
500 lAlsthrValLleAsnLysAspGlyLeuThrThrProLysAlaGlyAsn 516
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
787 TCATTTGCGCAGCTGCTCACCACATGTTTATCATATATGCCCCAAAGCA 836
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
517 Gly...GlyThrThrGlyThrAsnThrLleSer..... 526
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
837 AAATGGTTAATTAATGGGATATGCAACAGCAACCCCTATAT..G 883
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
527 .....ValThrLysAspGlyLleLysAlaGlyLysLysAlaLleThr 541
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
884 GAAAGCAATGCTCCAGCTAGCTGTAAGATGTTGTTCTATGATGAA 933
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
541 snValAlaSerGlyLeuThrGlyAla.....TyrAspAsp 551
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
934 ATCTTTGCTGAGATACCATTCAGTATTCAGCAACACATCAAAATG 983
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
552 Ala..... 552
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
984 GAATACTTTTAAACGACATTAATATGCGCGAGAAATTCGATGCCA 1033
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
553 .....AsnPheAspValLleAsnAsnSerAlaThrAspLeu..... 564
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1034 AACATAACACTATTCCTACCTTATAGATTAATAAACGACGAAACGCTTCA 1083
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
565 ..AsnArgHisValGluAspAlaTyrLys.....Gly 574
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1084 TTGTTTATATGTTCTTTATTCGAGACGACAGAGAACCTGTTAT.... 1128
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
575 LeuLeuAsnLysGluLysAsnAlaAsnLysGluProLeuValThrAs 591
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1129 ..CATGCTGAGGTGGGTCACAGTTATGACCCGACACTGATTAATGAG 1177
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
591 pSerThrAlaAlaThrValGlyAspLeuArgLys..... 602
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1178 AAAATATTTCTTTATGCAAGAAAGAAAGCAATGTTACTATCCAGC 1227
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
603 .....LeuGlyTyrValValSerThrLys 610
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1228 AACATCAACCAAGCGCGCGGCTTTGTTATTTGAGGTAATTTTACGT 1277
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

611 Asn..... glythrlyslucluserasn... glnva 620
 1278 CTGGCTAAAAACACGAAGACCTGGCAAGCGCGGCTTCATATCATGTG 1327
 620 LysGlnAlaAspGluValLeuPheThrGlyAlaGly..... 632
 1328 ATGCAGTACCGTACTTGGAAAGTAACGGCGGCGAAGCAAGCCGCTG 1377
 633 ..AlaLeuThrValThrSerLys..... 639
 1378 TCACAAATGCGCAAGGACGCGCTGGTTCACGCCAAGGGGAAACCA 1427
 640 SerGluAsnGlyLysHisThrIle..... 647
 1428 AGCGTGCACGCTGGCGGCGGTAAGTCATCTTAGATGACGACGGCG 1477
 648 ...ThrValSerValAlaGluThrLys..... 655
 1478 AGCATCAAGGCAAAAAACACAGCCTTAGTGAATCGGCTTGTCACGGC 1527
 656AlaAspCysGlyLeuGluLysAsp 663
 1528 AGGGGAGACGGTGCACATGATGCCGATATTCAGTTACACCCGACAACT 1577
 664 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnValIle 680
 1578 CTATTTCGCTTTGCGCGGCGGCGCTTGATTTGAACGGCGCATTCGCTT 1627
 680 uThrValGlyLysAsnGlyThrIleValIleThrLysGlyLysPheGluThrV 697
 1628 CGTTCACCGCATTCACAAATACCGATGAGAGGGCGATGATGTCAC... 1674
 697 alLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLysAsp 713
 1675CACAAATCAAGACAAAGAA..... 1692
 714 AlaThrIleAsnAspAlaAspLysLysValAlaIleThrValLysAspValAl 730
 1693TCCACCGTTACCATTCACAGGCATTAAGATATTTACTA 1729
 730 aThrAlaIleAsnSerAlaIleThrPheValLysThrGluAsnLeuThrT 747
 1730 CAAAC...GGCAATTAACAACACTTGATGCAAAAAAGAAATTCCTAC 1776
 747 hrSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAlaLeu 763
 1777 AACGGTTGG.....TTTGGCGAGAAAGATGCACCA 1808
 764 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysVally 780
 1809 AACGAAGGGGCGG.....CTCAATCTGATTAATACC 1837
 780 SarGAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuIVal 797
 1838 AACCGAGAGAGCGGATCGCACTTACTGCTTCCGCGGCAACAAATTTA 1887
 797 yThrAlaLysValSerAspThrLeuThrIleGlyLysThrProThr 813
 1888 AACGGC.....AATATCAAGCAACAAACGGCA 1916
 814 GlyGlyThrThrAlaThrProLysValAsnIleThrSerThrIleAspG 830
 1917 ACTGTTTTTCAGCGGACGACCGACCGCCTTACATCATTTAGGA 1966
 830 yLeuAsnPheAla...LysGluThrAlaAspAla.....SerGlyS 843
 1967 GCGGGGTGCACAAATGGAAGTATCCACACAGAGAAATCGTGGGAC 2016
 843 erLysAsnValIlyrLeuLysGlyIleAla..... 852
 2017 AACGATTTGATGACCGCATTTAAACGGGAAACTTCATATTCAG... 2064
 853 ThrThrLeuThrGluProSerAlaGlyAlaLysSerHisValAspIle 869
 2065 .GGCGAGAGGGNGGTTCCCGCAATGTTCGCAATGGA..... 2106
 869 uAsnValAspAlaThrLysLysSerAsnAlaIleSerIleGluAspVal 886
 2107GGCGATTTGCATTTAAGCAATCACCGCCCAACAGATTTTCGCTC 2151
 886 euArgAlaGlyTrpAsnIleGlnGlyAsnGlyAsnAsnValAspThrVal 902
 2152 GCACCGCATCAAGCCACACAAATCTGTACAGCTTCGGACTGGACGGTCT 2201
 903 AlaThrLysAsp.....ThrValAsnPheThrAspSerThrGlyThr 917
 2202 GACAAGTTGTACGAAAAAACCAATTCACCGAGTAA.....GTAGTG 2245
 917 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspVal 934
 2246 CTTCATTGACGACAGCGCATC..... 2268
 934 yslleGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 950
 2269AAGGCATATGCACCTTGCGCA 2291
 951 ThrGlyLysAspLeuLysAspAlaAsnGlyAlaThrValSerGluAs 967
 2292 TCACGCTCATTTAAATCTCACAGAGCTTGCACACTCAACGCAATCTTA 2341
 967 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIle 984
 2342 GTGCAGCGGAGACAGCGCACTATAGGTTAGC...CGCAACGCCACCCA 2388
 984 spAlaValAsnLysSerGlyTrpArgValThrGlyGluGlyAlaThrAla 1000
 2389 AACGGCAACCTCAGCCTGTG.....GGCAATGCCAAGCAACTTTAA 2432
 1001 GluThrGlyAlaThrAlaValAsnAlaGlyAsnAlaGluThrValThr 1017
 2433 TCAAGCCACATTA.....AACGGCAAC.....ACATCGG 2461
 1017 rGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThrAlaThrVal 1034
 2462 CTTGGCAATGCTTCATTTAATTCACCAACAGCGCGTACAAACGGC 2511
 1034 erLysAspAsnGlyAsnIleAsnValLysTrpAspValAsnValGlyAsp 1050
 2512 AGCTGACGCTTCGCAACAGCTTAAGCAACGTAAGCCATTCGCCACT 2561
 1051 GlyLeuLysIleGlyAspAspLysLysIleValAlaAspThrThrIle 1067
 2562 C.....AAGCGCAATGTCTCCCTA..... 2580
 1067 uThrValThrGlyLysLysValSerValProAlaGlyAlaAsnSerVal 1084
 2581GCCATTAAGCGATATTCATTTGAA 2607
 1084 snAsnAsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1100
 2608 AACAGCGGCTTACCGGAAAA..... 2628
 1101 AsnLeuSerTrpThrAlaLysAlaAspLysTyThrAlaAspGlyGluSerG 1117
 2629ATCAGCGCGGC..... 2640
 1117 uGlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLys 1134
 2641AAGGATACGGCATTTAC 2658
 1134 leGlyLysAsnLeuLysValLysLeuSerGluLysAspPheThrLysSer 1150
 2659 TTTAAAGACACGGAATGACGCTGCGCTG.....GGCAGCA 2696
 1151 leuGlnAspThrLeuThrGlyLysThrSerIleThrLeuGlyGlyThrAl 1167

```
2697 ATTAGCAAT.....TTAACTTGACAGCCACA 2728
      |||:|
1167 aasnlglyargasnphrghlythrvalileasnlyaspglyleuthrli 1184
      |||:|
2729 TTACACTCAATTCGGCTATCGACAGATCGCGCAGCGGCAACCGCG 2778
      |||:|
1184 lethrleu.....alaasnlglyalaalagly 1193
      |||:|
2779 AGTCGGCAGATGGCGCGCGCGCTTCCGCGCTTCCCTATATCCGT 2828
      |||:|
1194 thraspha.....serasnlyasnthrliaseriva 1204
      |||:|
2829 TACGCGCCCAACTTCGGCAGATCCGTTTCAACACGCTGACGGTAAACG 2878
      |||:|
1204 lthr.....lysaspglyileseralaglyasnlysgluilethrnsn 1219
      |||:|
2879 GCAATTAACGGTCACGAGAACATTCCGCTTATGTCCGAGACTCTCGGC 2928
      |||:|
1219 alyls.....seralaleuylthr 1225
      |||:|
2929 TACCGCAGCGCAATTAAGCTGGCGAAGTTCCGAAGCCACTTACAC 2978
      |||:|
1226 tyrllyasphthrghlnasnthrvalaspglythrghlnasplysgluphe 1242
      |||:|
2979 CTGGCTGTCAACAATACCGGCAC..GAACCGTAACTCTCGACCAAT 3025
      |||:|
1242 salalaivalyllyasnalaasnlglylupevalglylysasnlgly 1259
      |||:|
3026 TCAGCGTGTGAGAGAAAGACACACACCGCTGCCAAAATCTTAAT 3075
      |||:|
1259 lathrvalseralalythrashpn..... 1267
      |||:|
3076 TTCACCTTCGAACACACGCTGATGCCGCGCATGCGCTTACACT 3125
      |||:|
1267 ..... 1267
      |||:|
3126 TATCCGCAAGAGCGGAGTCCGCTGATTAATCCGTCAAAGAACAG 3175
      |||:|
1268 .....asnlglyshisthrvalthrileaspyalalaglyalal 1281
      |||:|
3176 AGCTTCCGACAAACTCGGCAAGCGGGAAGAAACAGAGCCGCTTGAGC 3225
      |||:|
1281 ysvalglyaspglyleuyllys.....AspThrAspglylystlely 1295
      |||:|
3226 GCAAAA.....CAGGCACA 3239
      |||:|
1296 leulysvalaspnthrashpnlyasnleuylthrvalasphalath 1312
      |||:|
3240 ACTTGGCGCAACACAGCGGAGAAAGACGCGCAAGCCTTGAGC 3289
      |||:|
1312 rlysglyalaserivalalylsglylupeasnalaivalthrthraspa 1329
      |||:|
3290 CGCTGATTCGGCGCGCGCAATCCACGCAAAAGCAGAACTGTTCG 3339
      |||:|
1329 lathrthrilainglythrashnalaasnlglylyvalalval 1345
      |||:|
3340 GAACCGCGCGCAGCGAGCGGGAAGAACCGCATATGACGCGGCA 3389
      |||:|
1346 lys.....glyserasn..glyalathralathrgl 1355
      |||:|
3390 GGAAGAAAAAAGCGGTGCGAGCGGTAAGACACCGCTTGCGCAAC 3439
      |||:|
1355 urthrashplylyvalalathrvalglyaspyalalalysalalea 1372
      |||:|
3440 AGCGCGAGCG.....GAACCGCGCGCTTACACCGCC 3474
      |||:|
1372 snasphalaalathrphvallyvalglusnaspsaspheralathrile 1388
      |||:|
3475 TTCCCCCGCGCGCGCGCGCGCGGATTTGGCGCAACGCGGCCCA 3524
      |||:|
1389 .....AspAspSerProthrashpspglyal 1397
      |||:|
3525 ACCGCAACCCCAACGCGAGCGCATGATCAGCGTTATGCCAATAGCG 3574
      |||:|

1397 aasnspalaleuylsalaglyAspThrleuthrleuylsalaglysa 1414
      |||:|
3575 GTTTGAGTAATTTTCCGCCACCTCAACAGCGTTTGGCGCTACAGAC 3624
      |||:|
1414 snleuylvallylsarargaspglylysasnliethrphleuilaasn 1430
      |||:|
3625 GAATTGACCGCGGTGTTCCGAAAGACCGCGCAACGCGCTTGGAACG 3674
      |||:|
1431 Aspleu.....servallyseral 1437
      |||:|
3675 CGGCATCCGGAGACCAACACTACCGTTCCGAAGATTTCCGCGCTACC 3724
      |||:|
1437 athrvalaserasp..... 1441
      |||:|
3725 GCCAACAACCGACCTCGCCCAATCGTATGCAGAAAACCTCGGACG 3774
      |||:|
1442 .....lysleuserleuylthrashnlglyasnlysalasnliethrser 1456
      |||:|
3775 GGGCGCGTGGCATCCTGTTTCG...CACAACCGAGCGCA..... 3813
      |||:|
1457 AspThrlysglyleuylspheralealysaspserythrghlyasphal 1473
      |||:|
3814 .AACACCTTCGACGACGCGCATCGCAACTCGGACGCGTTGCCACGCTG 3862
      |||:|
1473 aasnliethileuasnlglylealaserthrleuthr..... 1485
      |||:|
3863 CCGTTTCGGGCAATACGCGATCGGCGAGTTCGCATCGGCATCAGCGG 3912
      |||:|
1486 .....AspThrleuileuasnser 1491
      |||:|
3913 GCGCGCGTTTAACTAGCAGCGACGCTTTCAGACGCGATCAGCAAAAT 3962
      |||:|
1492 glylathrthrashnleugly.....glyasnlglylethrashpnlg 1506
      |||:|
3963 CCGCGCGCGCTGCTGCTATACGCGATTCAGGCAAGTACCGGAGGT 4012
      |||:|
1506 ulyslystrghlaala.....servallyaspyalaleuasnlaelyt 1521
      |||:|
4013 TCGGCGGATTCGGCATCAACCGCACATCGCGCACGCGCTATTTCGTC 4062
      |||:|
1521 tpsasnvalargglyalyspro..... 1528
      |||:|
4063 CAAAAGCGGATTCGATACGAAAGCTCAATATTCGCCACCGCGGCT 4112
      |||:|
1529 Alaseralaaasnlglyalgluasnliethasphavalala..... 1542
      |||:|
4113 TGCAATTCACCGCTACCGCGCGGCGCATTAAGCAGATTATTCATAAC 4162
      |||:|
1542 ..... 1542
      |||:|
4163 CGGCGCACACATTTTCATCACGCTTATTGAGCTGCTTATACCAT 4212
      |||:|
1543 .....ThrTyAspThrValasp 1548
      |||:|
4213 GCGCGTTCGGCAAAATCCGAAGCGCGCTCATATACCGCGTATGGCGCA 4262
      |||:|
1549 phevalserglyAspIysasphr...ThrservalthrIgluserly 1564
      |||:|
4263 GGAATTCGGCAAAACCGGAGTGGGAAATGGGCGCTTAACCGCAATC 4311
      |||:|
1564 saspsasnlglyshrglythrghlvalylslelyalalyslythrservali 1581
      |||:|
4312 .....AAAGTTTCACGCTGCTCCCTCACGCTGCGCC 4344
      |||:|
1581 lelyasphishasnlglylsleuphethr..... 1590
      |||:|
4345 GCCAAGGCGCGCAATTTGAACGACGACGACGCGGCGATC 4386
      |||:|
1591 .....Glylysgluileuylsasphalaasnasnlglyval 1602
      |||:|

seq_name: /cgn2_6/ptodata/1/laa/6B_COMB.pep:US-09-377-155-33
```

[illegible]


```
1278 CTCGCCTAAAAACACAGCAAGCTGGCAAGCGCGGCTTCATATCATGTG 1327
      ::::: ::::: |||||
620 llysglnAlaaspGluValLeuPheThrGlyAlaGly..... 632
1328 ATGGCAGTACCGTTACTTGGAAAGTAAAGCGCGTGGCAAAAGACCGCTG 1377
      ::::: ||||| |||
633 ..AlaIleThrValThrSerLys..... 639
1378 TCCAAATATGGCAAAAGCAGCGCTGCTGTTACAGCCAAAGGGAAACCA 1427
      ||||| |||||
640 SerGluAsnGlyLysHisThrIle..... 647
1428 AGGCTCGGTACGCTGGCGGCGAAGTAAAGTCATCTAGATCAGAGCGG 1477
      ::::: |||||
648 ...ThrValSerValAlaGluThrLys..... 655
1478 ACGATCAAGCAAAAAACACAGCTTTAGTGAATCGGCTGTGACGCGC 1527
      ::::: |||||
656 .....AlaAspCysGlyLeuGluLysAsp 663
1528 AGGGGAGCGGTGCACATGATCGGATATCATGTTCAACCCCGCAAACT 1577
      ||||| ::::: |||
664 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnValLe 680
1578 CATTTTCGCTTTTCGGCGGCGAGCTTTGATTTGAAGCGGATTCGCTTT 1627
      ||| ::::: |||
680 uThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluThrV 697
1628 CTTTCACCGCATTCACAAATACCGATGAAGGCGCATGATGTTCTCAAC... 1674
      ::::: |||||
697 alLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLysAsp 713
1675 .....CACATCAAGACAAGAA..... 1692
714 AlaThrAlaAsnAspAlaAspLysValAlaThrValLysAspValAl 730
1693 .....TCCACGCTTACCATTCAGCATTAAGATATTACTA 1729
      ||||| ::::: |||||
730 aThrAlaIleAsnSerAlaIleThrPheValLysThrGluAsnLeuThr 747
1730 CAAC...GCAATATAACACACTTGATGACAAAAGAAATTCCTTAC 1776
      ||||| ::::: |||
747 hSerIleAspGluAspAsnProThrAspAsnGlyLysAspAspAlaLeu 763
1777 AACGGTTG.....TTGGCGGAGAAAGATGCAACCA 1808
      ::::: |||||
764 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysVal 780
1809 AACGAGCGGCGG.....CTCAATCTGAATTACC 1837
      ||||| ::::: |||
780 sArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGluVal 797
1838 AACCGGAAGACGATCGCACTTACTCTTCCGCGGAAACAATTTA 1887
      ::::: |||||
797 yThrAlaLysValSerAspThrLeuThrIleGlyLysAsnThrProThr 813
1888 AACGCG.....AATTCACGCAAAACAACGCGCA 1916
      ::::: |||||
814 GlyGlyThrThrAlaThrProLysValAsnIleThrSerThrAlaAsp 830
1917 ACTGTTTTCAGCGGAGACCGACCGACCGCTTCAATCATTTAGGA 1966
      ||| |||||
830 yLeuAsnPheAla...LysGluThrAlaAspAla.....SerGlyS 843
1967 GCGGGGTGCTCAAAATGAAGTATCCCAAGAGCAAAATCGTGGGAC 2016
      ::::: |||||
843 eLysAsnValTyLeuLysGlyLeaAla..... 852
2017 AACGATTGGATCGACCGCACATTAAGCGGAAATCTCATATTAC... 2064
      ::::: |||||
853 ThrThrLeuThrGluProSerAlaGlyAlaLysSerHisValAspLe 869
2065 .GGCGGACAAGCGGTGGTTCCGCAATGTTCGCAAAAGTGAA..... 2106
      ::::: |||||

      ::::: |||||
869 uAsnValAspAlaThrLysLysSerAsnAlaIleSerIleGluAspVal 886
2107 ....GCCGATTGGCATTTAAGCATACAGCCCAACGACCTTTGGGTGC 2151
      ::::: |||||
886 euArgAlaGlyTrpAsnIleGlnGlyAsnGlyAsnAsnValAspTrpVal 902
2152 GCACCCATCAACAGCCACACAACTCTGTACAGCTTGGAGTGGAGGCTCT 2201
      ||| ::::: |||||
903 AlaThrTrpAsp....ThrValAsnPheThrAspAspSerThrGly 917
2202 GACAACTGTGACGAAACCAATTCACGACGATAA.....GTCATTG 2245
      ||| ::::: |||
917 rThrThrValThrValThrGluLysAlaAspGlyLysGlyAlaAspVal 934
2246 CTTCAATTGACACAGACCGCATC..... 2268
      ::::: |||||
934 ySileGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 950
2269 .....AGAGCAATGTGAGCCTTGCGCA 2291
      ::::: |||||
951 ThrGlyLysAspLeuLysAspAlaAsnAsnGlyAlaThrValSerGluAs 967
2292 TCACGCTCATTTAAATCTCACAGACTTGCACACACTCAACGCAATCTA 2341
      ::::: |||||
967 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIleA 984
2342 GTGAGCGCGGAGACACGACCTATACGTTAGC...CGCAAGCGCACCA 2388
      ::::: |||||
984 sPalaValAsnLysSerGlyTrpArgValThrGlyGluGlyAlaThrAla 1000
2389 AACGCAACCTCAGCCTCTGTC.....GGCAATGCCCAACCAATTTAA 2432
      ::::: |||||
1001 GluThrGlyAlaThrAlaValAlaAsnAlaGlyAlaAsnAlaGluThrVal 1017
2433 TCAAGCCACATTA.....AACGCAAC.....ACATCGG 2461
      ::::: |||||
1017 rGlyThrSerValAsnPheLysAsnGlyAsnAlaThrThrAlaThrValS 1034
2462 CTTGGCAATACGCTTCATTTAATCTAAGCAACAGCGGTACAAAGCGC 2511
      ::::: |||||
1034 eLysAspAsnGlyAsnIleAsnValLysTrpAspValAsnValGlyAsp 1050
2512 AGTCTGACGCTTTCGACACAGCTAAGCAACGTAAGCAATTCGCGACT 2561
      ::::: |||||
1051 GlyLeuLysIleGlyAspAspLysLysIleValAlaAspThrThrIle 1067
2562 C.....AACGCAATGTCTCCCTA..... 2580
      ::::: |||||
1067 uThrValThrGlyGlyLysValSerValProAlaGlyAlaAsnSerValA 1084
2581 .....GCCGATAGCGCATTCATTCATTTGAA 2607
      ::::: |||||
1084 sAsnAsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1100
2608 AACACGCGCTTTCACGGA..... 2628
      ::::: |||||
1101 AsnLeuSerTrpThrAlaLysAlaAspLysTrpAlaAspGlyLysSerG 1117
2629 .....ATCAGCGGCGG..... 2640
      ::::: |||||
1117 uGlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysA 1134
2641 .....AACGATACGCAATTACAC 2658
      ::::: |||||
1134 lAglyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTySer 1150
2659 TTTAAAGACGCGAATGGACGCTGCGCTG.....GGACGCA 2656
      ::::: |||||
1151 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAl 1167
2697 ATTAGGCAT.....TTAACTTGACAACGCGACCA 2728
      ::::: |||||
```

```

1167 aasnlglyarqasnsprthrglythrvalleasnlyaspglyleuthri 1184
2729 TTACACTCAATTCGCCCTATCGACACGATGGCGGCGGCAACCGCG 2778
1184 lethrleu.....Alaasnlglyalalaalagly 1193
2779 AGTGGCGAGATGCGCGCGCGCGCGCTTCCTTATTCCTTATTCCTT 2828
1194 Thrspala.....Serasnlglysnthrleu 1204
2829 TACGCGCGCACTTCGGCAGATCCCTTTCACACGCTACCGGTAAAG 2878
1204 lthr.....LysaspglyleuSerlalaalasnlysglyleuThrAsp 1219
2879 GCAATTTGAAGGTCAGGGAACATTCGCTTATGTCGGAATCTTCGCG 2928
1219 allys.....Serlalaalasnlythr 1225
2929 TACGCGCGCGCAATTTGAAGCTGCGGAAAGTTCCGAAAGCACTTAC 2978
1226 TyllysasprthrGlnasnthrAlaaspGluThrGlnasplsgluheh 1242
2979 CTGGCTGTCAACAATACCGCAAC...GAACCGTAACTTCGAGCAT 3025
1242 salalavalalysasnlaasnlglyalasphevalglylysasnlgly 1259
3026 TGACGCTAGTGGAAAGAAAGACACACACCGCTGCGAAATCTTAAT 3075
1259 lathrvalSerlalaalysThrAspsn..... 1267
3076 TTCACCTCGAAAGCAACGACGTCGATGCGCGCATGCGGTATTCAGCT 3125
1267 ..... 1267
3126 TATCCGAAAGCGCGGATTCGCTGCATATCCGTCAAAGCAACAG 3175
1268 .....Asnlglyshthrvalthrleasvalalagluvalal 1281
3176 AGCTTCGCAACAATCGCGAGCGGGAAGAAAGAGCGCCCTTGACG 3225
1281 ysvalglyaspglyleuGluLys.....AspThrAspglylysleuLys 1295
3226 GCAAAA.....CAGGCACA 3239
1296 leulysvalaspasnthrAspglyasnleulthrvalaspAlaThr 1312
3240 ACTTGGCGCGCAACACAGCGGAAAGACACCGCAACCTTGACG 3289
1312 rlysglyalaservalalalysglylupheasnalaValThrThrAspa 1329
3290 CGCTGATTCGCGCGCGCATGCGACGCAAGAAAGCAAGAAAGTGTGCC 3339
1329 lethrthrAlaGlnGlythrAsnAlaasnlglyalasphevalVal 1345
3340 GAACCGCGCGCGAGCGGGAATAATGCGCATTTATGACGCGGA 3389
1346 Lys.....glyserasn...glyAlaThrAlaThrGln 1355
3390 GGAAGAGAAAGAAAGCGGTGCGAGCGGATTAAGACACCGCTTGCGCAAC 3439
1355 uthrAspglyLysLysValAlaThrValglyAspValAlaLysAlaIle 1372
3440 AGCGGGAAGCG.....GAACCGCGCGCGGTACACACCGCC 3474
1372 snAspAlaAlaThrphValLysValGluasnAspSerThrAlaThrIle 1388
3475 TTCGCCCGCGCGCGCGCGCGCGGATTCGCGCAACCGCGCGCCA 3524
1389 .....AspAspSerProthrAspAspglyAl 1397
3525 ACCGCAACCCCAACCGCGCGCGCATGATGACCGTATGCGCAATGCG 3574
1397 aasnAspAlaLeuLysAlaGlyAspThrLeuThrLeuLysAlaGlyLysa 1414
3575 GTTTGAGTGAATTTCCGCCACGCTCAACAGCGTTTTCGCGTACAGAC 3624
1414 snleuLysValLysArgAspglyLysasnlethrphAlaLeuAlaasn 1430
3625 GAATTGACCGCGGTGTTGCGGAAGACCGCGCAACCGCGTTTGCAAG 3674
1431 Aspleu.....SerValLysSerAl 1437
3675 CGGCATCCGGGACACCAACACTACGTTTCGAAGATTCGCGGCTAC 3724
1437 aThrValSerAsp..... 1441
3725 GCCAACAACCGACCTGCGCAATCGTATGCAAGAAACCTGCGACG 3774
1442 .....LysleuSerleuGlyThrAsnlglyAsnlyValasnlethrSer 1456
3775 GCGCGCGTCGCAATCGTTTTCG...CAACCGGACCGGA..... 3813
1457 AspThrLysGlyleuasnphAlaLysAspSerLysThrGlyAspAspAl 1473
3814 .AACCTTCGACGACGCGATCGGCACTGCGCGCGCTTGCCACGCTG 3862
1473 aasnleuHisleuasnlglylealaserThrleuThr..... 1485
3863 CCGTTTCGCGCAATACGGCATCGGCGATGACATCGGCGATCGCGCG 3912
1486 .....AspThrleuLeuAsnSer 1491
3913 GCGCGCGGTTTATGACGCGAGCGCTTCGACGCGCATGAGCAAAAT 3962
1492 GlyAlaThrThrAsnleuGly.....GlyasnlglyleThrAspAsnG 1506
3963 CGCGCGCGCGTCTGCTGATTCAGCGATTCAGCAAGATACCGCGAGTT 4012
1506 uLysLysAlaGlyAlaAla.....SerValLysAspValleuasnlaGlyT 1521
4013 TCGGCGATTCGCGCATCGAACCGCACATCGCGCGCAACGCTATTCTC 4062
1521 rpsnValArGlyValLysPro..... 1528
4063 CAAAAGCGGATTCGCGATGCAAGAAAGTCAATATCGCGCGCGCT 4112
1529 AlaSerAlaAsnAsnGlnValGluasnleuAspHevalAla..... 1542
4113 TGCATTCACCGCTACCGCGCGCGCATTAAGCAGATTATTCATTCAAC 4162
1542 ..... 1542
4163 CGGCGCACACATTCATTCACGCGCTTATTTGAGCTGTCTATACCAT 4212
1543 .....ThrTyrAspThrValAsp 1548
4213 GCGCGTTCGCGCAAGTCCGAACGCGCTCAATACCGCGATTGCGCA 4262
1549 PheValSerGlyAspLysAspThr...ThrSerValThrValGluSerly 1564
4263 GGATTCGGCAAAACCCGACAGTGGGAATGGGCGTAAAGCGCAATC. 4311
1564 saspsnlglyLysArgThrGlnValLysIleGlyAlaLysThrSerVal 1581
4312 .....AAAGTTTCACGCTGTCTCCATCAACGCTGCGCGC 4344
1581 leLysAspHisAsnlglyLysleuPheThr..... 1590
4345 GCCAAGGCGCGCAATTTGAAGCGCAGCAGACGCGCGGCAATC 4386
1591 .....GlyLysGlyleuLysAspAlaAsnAsnleuVal 1602

```

seq_name: /cgn2_6/prodata/1/laa/bb_comb.pep:us-08-913-942-4
seq_documentation_block:
; Sequence 4, Application US/08913942


```

1034 AACATCAACACTATCTCTACCTTATAGATTAAAAACGACGACCGTTCAA 1083
      :::::||||| ::::| ::::|
565 ..AsnArgHisValGlnAspAlaTrpLys.....Gly 574
1084 TTGTTTAATGTTTCTTTATCCGAGACAGACGACGACCTGTTTAT..... 1128
      :::::||||| :::::|||||
575 LeuLeuAsnLeuAsnGlnLysAsnAlaAsnLysGlnProLeuValThrAs 591
1129 .CATGCTGAGGTGGGTCAACAGTTATCCGACGACGACGACGACGACGAG 1177
      :::::||||| :::::|||||
591 pSerThrAlaAlaThrValGlyAspLeuValArgLys..... 602
1178 AAAATATTTCTTTATTGACAAAGAAAGGTAATGATATTACTTACGACG 1227
      ||| :::::|||||
603 .....LeuGlyTrpValValSerThrLys 610
1228 AACATCAACCAAGCGCGCGGCTTTGATTTGAGGTAATTTTACGCT 1277
      ||| :::::|||||
611 Asn.....GlyThrLysGlnLysSerAsn...GlnVal 620
1278 CTCGCTTAATAAACAACGAAACGAGTGGCAGCGCGGCTTCATATACAGTG 1327
      ||::: :::::|||||
620 LysGlnAlaAspGlnValLeuPheThrGlyAlaGly..... 632
1328 ATGGCAGTACCGCTTACTTGAAAGTAAACGCGTGGCAACGACCGCTG 1377
      :::::|||||
633 ..AlaAlaThrValThrSerLys..... 639
1378 TCCAAATTCGGCAAGGCGACGCTGCTGCTTCAACCCAAAGGGAACCA 1427
      ||::: |||||
640 SerGlnAsnGlnLysHisThrIle..... 647
1428 AGGCTCGTACGCGTGGCGACGCTAAAGTCATCTTATGATCAGACGCGG 1477
      :::::|||||
648 ....ThrValSerValAlaGlnLys..... 655
1478 ACGATCAAGGCAAAAAACGCTTATGTAATCGCTTGTGACGCGC 1527
      :::::|||||
656 .....AlaAspLysGlnLysLeuLysAsp 663
1528 AGGGGACGCTGCACTGAATGCCGATTAAGTTCATCAACCCGACCAACT 1577
      ||:::|||||
664 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnValLe 680
1578 CTATTTCGCTTTCGCGCGGACGCTTGGATTGAAACGCGCATTCGCTTT 1627
      ||| :::::|||||
680 uThrValGlyAsnAsnGlnYThrAlaValThrLysGlnLysPheGlnThr 697
1628 CGTTCACGCGCATTCAAATACCGATGAAGGCGCATGATGTTGTCAC... 1674
697 allysThrGlnLysAlaThrAspAlaAspArgGlyLysValThrValLysAsp 713
1675 .....CACATCAAGCAAAAGA..... 1692
714 AlaThrAlaAsnAspAlaAspLysLysValAlaAlaThrValLysAspValAl 730
1693 .....TCACCGCTTACCATTAACGCAATAAAGATATTACTA 1729
730 aThrAlaIleAsnSerAlaAlaThrPheValLysThrGlnLysLeuThr 747
1730 CAACC...GGCAATACAACTGGATTACCAAAAAAGAAATGTCCTAC 1776
      ||:::|||||
747 HisSerIleAspGlnAspAsnProThrAspAsnGlnLysAspAspAlaLeu 763
1777 AACGCTTG.....TTGCGAGAAAGATGACACCA 1808
764 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysValLys 780
1809 AACGACGCGCG.....CTCAATCTGATTAAC 1837
780 sArgAspGlnLysAsnIleThrPheAspLeuAlaLysAsnLeuGlnVal 797
1838 AACGGGAAGACGGATGCACTTACTGCTTCCGCGGACAAATTTA 1887
      :::::|||||
797 yThrAlaLysValSerAspThrLeuThrIleGlyLysAsnThrProThr 813
1888 AACGCG.....AATATCAAGCAAAACGACCA 1916
      :::::|||||
814 GlyGlyThrThrAlaThrProLysValAsnIleThrSerThrAlaAspGln 830
1917 ACTGTTTTCAGCGGACGACCGACCGGCTTCATCAATCATTTAGGA 1966
      ||| |||||
830 yLeuAsnPheAla...LysGlnThrAlaAspAla.....SerGlyS 843
1967 GCGGCTGTCAAAATGAAAGTATCCACAGACGAGAATCGTGGGAC 2016
      ||| :::::|||||
843 erLysAsnValTrpLeuLysGlyIleAla..... 852
2017 AACGATTTGATTCGACCGCATTTAAACGGGAACCTTCATATTCAG... 2064
      :::::|||||
853 ThrThrLeuThrGlnProSerAlaGlyAlaLysSerThrHisValAspLe 869
2065 .GCGGACAAAGCGGTGTTCCGCAATGTTGCCAAAGTGA..... 2106
      :::::|||||
869 uAsnValAspAlaThrLysLysSerAsnAlaAlaSerIleGlnAspValL 886
2107 ....GGCGATTTGCAATTTAGCAATCAGCCCAAGCATTTCGGTGC 2151
      :::::|||||
886 euArgAlaGlyTrpAsnIleGlnLysAsnGlnLysAsnValAspTrpVal 902
2152 GCACCGCATCAAAAGCCACACATCTGTACACGTTGACGAGTGGAGGCT 2201
      :::::|||||
903 AlaThrTrpAsp....ThrValAsnPheThrAspSerThrGlyTh 917
2202 GACAAGTTTACCGAAAAAACATTACGACGATAA.....GTGATG 2245
      ||| |||||
917 rThrThrValThrValThrGlnLysAlaAspGlnLysGlyAlaAspValL 934
2246 CTTCATTGACGACGACCGACATC..... 2268
934 ystIleGlyAlaLysThrSerValIleLysAspHisAsnGlnLysLeuPhe 950
2269 .....AGAGCAATGTACACTTGGCCA 2291
951 ThrGlyLysAspLeuLysAspAlaAsnGlnLysAlaThrValSerGlnAs 967
2292 TCACGCTCTTAATCTACAGACTTCCACACTCAGCGCACTTTA 2341
      :::::|||||
967 pAspGlyLysAspThrGlnLysLeuValThrAlaLysThrValIleAla 984
2342 GTTCAGGCGGAGACGACCTATACGCTTAGC...CGCAACGCGCACCA 2388
      :::::|||||
984 spAlaValAsnLysSerGlnTrpArgValThrGlyGlnGlyAlaThrAla 1000
2389 AACGCGACCTCAGCCTCGTG.....GGCAATGCCCAAGCACTTTAA 2432
      :::::|||||
1001 GlnThrGlyAlaThrAlaValAlaAsnAlaGlnLysAlaGlnLysThrVal 1017
2433 TCAGCGCACATTA.....AACGCGAC.....ACATCGG 2461
1017 rGlyThrSerValAsnPheLysAsnGlnLysAlaThrThrAlaThrValS 1034
2462 CTTCGACAAATGCTTCATTAATCTAAGCAACAGCGCGTACAAAGCGC 2511
      :::::|||||
1034 erLysAspAsnGlnLysAsnIleAsnValLysTrpAspValAsnValGlyAsp 1050
2512 AGCTGACGCTTTCGACACAGCTTAAGCAAAAGCAATTCGCGACT 2561
      :::::|||||
1051 GlyLeuLysIleGlyAspAspLysLysIleValAlaAspThrThrThrIle 1067
2562 C.....AACGCAATGTCCTCA..... 2580
1067 uThrValThrGlnLysLysValSerValProAlaGlyAlaAsnSerValAla 1084
2581 .....GCCGATAAGCGCATTAATTCATTTGAA 2607

```

```
1084 snaasnaLysLysLeuValAsnAlaGluGluLeuAlaThrAlaLeuAsn 1100
2608 AACAGCCGCTTACCGGAAAA..... 2628
1101 AsnLeuSerThrAlaLysAlaAspLysThrAlaAspGlyIleuSerG1 1117
2629 .....ATCAGCGCGC..... 2640
1117 uGlyLThrAspGlnGluValLysAlaGlyAspLysValThrPheLysA 1134
2641 .....AAGGATACGGCATTCAC 2658
1134 laGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrTyrSer 1150
2659 TTAAGAAGACGCAATGAGCGCTGCGTCG.....GSCACGGA 2696
1151 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyIleThrAl 1167
2697 ATTAGCAAT.....TTAACCCTTGACACGCCACCA 2728
1167 aasnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1184
2729 TTACAGCTCAATTCGCTATGACACGATGCGCGCGCAAAACCGCGC 2778
1184 IeThrLeu.....AlaasnGlyAlaAlaAlaGly 1193
2779 AGTGGCGAGATGCGCGCGCGCGCTGCGCGCTTCCTATTATTCCT 2828
1194 ThrAspAla.....SerasnGlyAsnThrIleSerVal 1204
2829 TACGCGCCCACTTGGCAGATCCGTTTCAACACGCTGACGCTAAG 2878
1204 IThr.....LysAspGlyIleSerAlaGlyAsnLysGluIleThrAsnV 1219
2879 GCAATATGACGCTCAGGAGACATTCGCTTATGTCGAACTCTTCGCG 2928
1219 alLys.....SerAlaLeuLysThr 1225
2929 TACCGCAGCGCAATTTGAAGCTGGCGGAAATTCGGAAGACCTTAC 2978
1226 TyrLysAspThrGlnAsnThrAlaAspGluThrGlnAspLysGluPheH 1242
2979 CTGGGCTGACACATACCGCGAAC...GAACCGTAACTTCGACCAAT 3025
1242 salAlaValLysAsnAlaAsnGluValGluPheValGlyLysAsnGlyA 1259
3026 TGACGGTAGTGAAGAAAAAGACACACACCGCTGTCGAAATCTTAAT 3075
1259 laThrValSerAlaLysThrAspAsn..... 1267
3076 TTCAACCTGCAAAACACACGCTGATGCCGCGCATGGCGTTATCAGCT 3125
1267 ..... 1267
3126 TATCCGAAAGACGCGGATTCGCTGCATATCCGTCAAAGAACAG 3175
1268 .....AsnGlyLysHisThrValThrIleAspValAlaGluAlaH 1281
3176 AGCTTTCGCAAACTCGCAGCGGAGGAAACACAGCGCGCTTGAGC 3225
1281 ysaValGlyAspGlyLeuGluLys.....AspThrAspGlyLysIleLys 1295
3226 GCAAAA.....CAGCGACA 3239
1296 LeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaTh 1312
3240 ACTTCCGCAACACACGCGGAAAAAGACACGCGCAAAAGCTTCAGC 3289
1312 rLyGlyAlaSerValAlaLysGlyIlePheAsnAlaValThrThrAspA 1329
3290 CGCTGATGGCGCGCGCATGCAACGAAAGGACAAAGCTTGGCC 3339
1329 laThrAlaGlnGlyThrAsnAlaAsnGluArgGlyLysValVal 1345
3340 GAACCGCGCGCGCAGCAGCGGGAATGCCGATATATCAGCGCA 3389
1346 Lys.....GlySerAsn...GlyAlaThrAlaThrG1 1355
3390 GGAAGAGAAAAACGGGTGACGCGATTAAGACACCGCTTGGCAAC 3439
1355 uThrAspLysLysValAlaThrValGlyAspValAlaLysAlaIleA 1372
3440 AGCGCGACAGC.....GAACCGCGCGCTACACCGCC 3474
1372 snAspAlaIaThrPheValLysValGluAsnAspSerIaThrIle 1388
3475 TTCGCCGCGCGCGCGCGCGCGCGGATTCGCAACCGACGCCCA 3524
1389 .....AspAspSerProThrAspAspLysAl 1397
3525 ACCGCAACCCCAACCGCAGCGCATGATCAGCCGTTATGCCAATAGC 3574
1397 aasnAspAlaLeuLysAlaGlyAspThrLeuThrLeuLysAlaGlyLysA 1414
3575 GTTTGAGTGAATTTCCGCCAGCCTCAACACGCTTTCGCGCTACAGC 3624
1414 snLeuLysValLysArgAspGlyLysAsnIleThrPheAlaLeuAlaAsn 1430
3625 GAATTGACGCGGTTGTCGGAAGACCGCGCGCAACGCGCTTGGACAG 3674
1431 AspLeu.....SerValLysSerAl 1437
3675 CGGCATCGCGGACACCAACACTACGCTTCGCAAGATTTCCGCGCTAC 3724
1437 aThrValSerAsp..... 1441
3725 GCCAACAACCGACCTGCGCCAAATCGTATGCAAAAACTCGCGAGC 3774
1442 .....LysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIleThrSer 1456
3775 GGGCGCGTGGCATCTGTTTCG...CACACCGGACCGGA..... 3813
1457 AspThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAspAl 1473
3814 AACACCTTCAGACGACGATCGGCACACTCGGCAACGCTTGGCCAGCTG 3862
1473 aasnIleHisLeuAsnGlyIleAlaSerThrLeuThr..... 1485
3863 CGGTTTGGGCAATACGCGATCGCAGGTTTCGACATCGGATCAGCGCG 3912
1486 .....AspThrLeuLeuAsnSer 1491
3913 GGGCGGCTTTAGTAGCGGACCTTTCAGACGCGATCAGAGCAAAAT 3962
1492 GlyAlaThrThrAsnLeuGly.....GlyAsnGlyIleThrAspAsnG1 1506
3963 CCGCGCGCGGCGGTGATTAAGCATTCAGGCAAGATACCGCGAGGT 4012
1506 uLysLysArgAlaAla.....SerValLysAspValLeuAsnAlaGlyT 1521
4013 TCGCGGATTCGCGATTCGACACCGCATCGGCGCAACGCGCTATTTCGTC 4062
1521 rPAsnValArgLysValLysPro..... 1528
4063 CAAAAAGCGGATTCAGATACGAAACGTCATATCGCACCGCGGCGCT 4112
1529 AlaSerAlaAsnGlnValGluAsnIleAspPheValAla..... 1542
4113 TCGATTCACCGCTACCGCGCGGCGATTAAGCAGATTATTCATCAAC 4162
1542 ..... 1542
4163 CGCGCAACACTTTCATACGCGCTTATTTAGCTGTCCTATACCAT 4212
1543 .....ThrTyrAspThrValAsp 1548
```

```
4213 GCCGCTCCGCAAGTCCGACGGCGTCATATCCGCCGTATTGGCGCA 4262
      |||||
1549 PheValSerGlyAspLysAspThr...ThrSerValThrValGluSerI 1564
      |||||
4263 GGATTTCGCAAAACCCGAGTGGGGAATGGCGGTAAACCGCAATC 4311
      |||||
1564 SASPAsnLysArgThrGluValLysIleGlyAlaLysThrSerValI 1581
      |||||
4312 .....AAAGTTTCACGCTGTCCCTCCACGGCGCCGCC 4344
      |||||
1581 IelYAspHisAsnGlyLysLeuPheThr..... 1590
      |||||
4345 GCCAAGGGCGCCCAATTGGAAGCCGACACAGCGCGGCGATC 4386
      |||||
1591 .....GlyLysGluLeuLysAspAlaAsnAsnGlyVal 1602

seq_name: /cgn2_6/protdata/1/1aa/6B_COMB.pep:US-09-669-974-33

seq_documentation_block:
/ Sequence 33, Application US/09669974
/ Patent No. 6333173
/ GENERAL INFORMATION:
/ APPLICANT: PEAK, Ian Richard Anselm
/ APPLICANT: JENNINGS, Michael Paul
/ APPLICANT: MOXON, E. Richard
/ TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
/ FILE REFERENCE: 065064/0128
/ CURRENT APPLICATION NUMBER: US/09/669, 974
/ PRIOR FILING DATE: 2000-09-26
/ PRIOR APPLICATION NUMBER: US 09/377, 155
/ PRIOR FILING DATE: 1999-08-19
/ PRIOR APPLICATION NUMBER: PCT/AU98/01031
/ PRIOR FILING DATE: 1998-12-14
/ PRIOR APPLICATION NUMBER: GB 9726398.2
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: Patentl Ver. 2.0
/ SEQ ID NO 33
/ LENGTH: 2353
/ TYPE: PRT
/ ORGANISM: Haemophilus influenzae
US-09-669-974-33

alignment_scores:
Quality: 253.50 Length: 1664
Ratio: 0.344 Gaps: 75
Percent Similarity: 44.231 Percent Identity: 19.591

alignment_block:
US-09-303-518D-653 x US-09-669-974-33 ..
Align seg 1/1 to: US-09-669-974-33 from: 1 to: 2353

7 AGAACCGCAACGCAACCGCA.....ACACCGCAAAAGCCCC 47
||| |||||
246 ThrGlyAspLysAsnThrLeuAspValValLeuThrAlaLysGluAsnG 262
||| |||||
48 TAAACCGCGCGCATCGCTTCTGCGCGCTTACTAGCCATATGCTGT 97
||||| |||||
262 YLysThrThrGluValLysPheThrProLysThrSerValLysGluI 279
||| |||||
98 CGTTCGCAATTCGCGCCACAGCGCGGCGGACACTTATTTGGCATC 147
||| |||
279 YAspGlyLysLeuPheThrGlyLysGluAsnAsnAspThrAsnLysVal 295
||| |||
148 AACTACCAATACTATCGGACTTTGCGCAAAATAAAGCAAGTTGGCA 197
||| |||
296 ThrSerAsnThrAlaThrAspAsnThrAspGluLysAsnGlyLeuVal 312
||| |||
198 CGGGCGCAAGATATTGGGTTTACACAAAAAAGGAGATTGTCGCGCA 247
||| |||
```



```

1084 sAsnAsnLysLysLeuValAsnAlaGluGlyLeuAlaThrAlaLeuAsn 1100
2608 AACAGCCGCTTTACCGGAAAA..... 2628
1101 AsnLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspLysGluSerG 1117
2629 .....ATCAGCGCGGC..... 2640
1117 uGlyGluThrAspGlnGluValLysAlaGlyAspLysValThrPheLysA 1134
2641 .....AAGGATAGCGATTACAC 2658
1134 laGlyLysAsnLeuLysValLysGlnSerGluLysAspPheThrSer 1150
2659 TTAAAGACAGCAGCATGAGCGCTGCGCTCG.....GGCAGCGA 2696
1151 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyGlyThrAl 1167
2697 ATTAGCCAAAT.....TTAAACCTTGACAAAGCCACCA 2728
1167 aAsnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1184
2729 TTACACTCAATTCGGCTATCGACAGANGCGGACAGCGCGCAAAACGGC 2778
1184 leThrLeu.....AlaAsnGlyAlaAlaAlaGly 1193
2779 AGTGGCGCAGATGCGCGCGCGCGCTTGGCCGCTTCCCTATATTCGT 2828
1194 ThrAspAla.....SerAsnGlyAsnThrIleSerVal 1204
2829 TACCGCGCCAACTTCGCGCAAAATCCCGTTCAACAGCGTGACGGTAACG 2878
1204 lThr.....LysAspGlyIleSerAlaLysAsnLysGluIleThrAsn 1219
2879 GCAAATTAAGCGTGACGAGACATTCGCTTATCTCGGAATCTTCGCG 2928
1219 aLys.....SerAlaLeuLysThr 1225
2929 TACCGAGCGCAAAATTGAAGCTGCGCGAAATTCGCAAGCAGCACTTAC 2978
1226 TyrLysAspThrGlnAsnThrAlaAspGluThrGlnAspLysGluPheH 1242
2979 CTGGCTGTCAACATACCGGCAC...GAACCCCTAAGTCTCGAGCAAT 3025
1242 sAlaAlaValLysAsnAlaAsnGluValGluPheValGlyLysAsnGlyA 1259
3026 TGACGCTAGTGAAGAGAAAGACAACACACGCTGTCCGAAATCTTAAT 3075
1259 laThrValSerAlaLysThrAspAsn..... 1267
3076 TTCACCTCGCAAAACGACAGCTGATCGCGGCATGCGCTTACAGCT 3125
1267 ..... 1267
3126 TATCCGCAAAAGCGGAGTTCCGCTGCATATCCGTCAAAGACAAG 3175
1268 .....AsnGlyLysHISThrValThrIleAspValAlaGluAla 1281
3176 AGCTTCGCAAAACTCGGCAAGCGGGAGACAACAGAGCGGCTTGACG 3225
1281 ySValGlyAspGlyLeuGluLys.....AspThrAspGlyLysIleLys 1295
3226 GCAAAA.....CAGGCACA 3239
1296 LeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaTh 1312
3240 ACTTCCGCGCAACACAGCGGAAAAAGACAACGCAAGCCTTGACG 3289
1312 rLysGlyLysSerValAlaLysGlyLupheAsnAlaValThrThrAspA 1329
3290 CGGATTTGCGCGCGCGCAATGCCAGCAAAAGCGCAAAAGTGTGGC 3339
||| ||| |||||:||||:||||: |||:|
1329 laThrThrAlaGlnGlyThrAsnAlaAsnGluArgLysValValVal 1345
3340 GAACCGCGCGCGGACGAGCGGGAATGCCGCTATTATGACGGCGA 3389
1346 Lys.....GlySerAsn...GlyAlaThrAlaThrGlu 1355
3390 GGAAGAGAAAAACGGGTGACGCGATTAAGACACACCGCTTGGCGAAC 3439
1355 uThrAspLysLysLysValAlaThrValGlyAspValAlaLysAlaIleA 1372
3440 AGCGCGAAGCG.....GAAACCGCGCGGCTACACCGCC 3474
1372 snAspAlaAlaThrPheValLysValGluAsnAspSerAlaThrIle 1388
3475 TTCCCGCGCGCGCGCGCGCGCGGATTTCCGCGCAACCGCACCCA 3524
1389 .....AspAspSerProThrAspAspGlyAl 1397
3525 ACCGCAACCCCAACCGCAGCGGACCTGATACCGCTTATGCAATPACG 3574
1397 aAsnAspAlaLeuLysAlaGlyAspThrLeuThrLeuLysAlaGlyLysA 1414
3575 GTTTGAGTGAATTTCCGCGACGCTCAGCGCTTTCGCGTACAGGAC 3624
1414 snLeuLysValLysArgAspGlyLysAsnIleThrPheAlaLeuAlaAsn 1430
3625 GAATTGACCGCGGTGTTTTCGGAAGACGCGCGCACGCGCTTTGGACAG 3674
1431 AspLeu.....SerValLysSerAl 1437
3675 CGGCATCCGGGACACCAACACTACCTGTCGAGAGATTTCCGCGCTTAC 3724
1437 aThrValSerAsp..... 1441
3725 GCGAACAACCGCACTCGCGCAATCGGTATGACGAAACCTTGAGAC 3774
1442 ....LysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIleThrSer 1456
3775 GGGCGGTCGCGCATCTGTTTCG...CACACCGGACCGGA..... 3813
1457 AspThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAspAspAl 1473
3814 .AACACTTCGACGACGCGCATCGCACTCGGACGCGCTGCGCACGCTG 3862
1473 aAsnIleHisLeuAsnGlyIleAlaSerThrLeuThr..... 1485
3863 CCGTTTGGCGCAATACGGCATTCGCGAGTTCGACATCGGCATACGCG 3912
1486 .....AspThrLeuLeuAsnSer 1491
3913 GCGCGGCTTTTATAGCGGACGCTTTCAGACGCGATTCAGAGCAAAAT 3962
1492 GlyAlaThrThrAsnLeuGly.....GlyAsnGlyIleThrAspAsnGlu 1506
3963 CCGCGCGCGCTGCTGCAATTACGGCATTCAGCAAGATATACCGGCAAGT 4012
1506 uLysLysArgAlaAla.....SerValLysAspValLeuAsnAlaGlyT 1521
4013 TCGGCGATTCGGCATGCAACCGCACATTCGGCGCAAGCGCTATTTCGTC 4062
1521 rPAsnValArgGlyValLysPro..... 1528
4063 CAAAAAGCGATTAACGATACGAAACGTAATATGCCACCGCGGCGCT 4112
1529 AlaSerAlaAsnAsnGluValGluAsnIleAspPheValAla..... 1542
4113 TGCAATTCAACCGCTACCGCGCGGCAATTAAAGCAGATTATTCATTCAAC 4162
1542 ..... 1542
4163 CGGCGCAACACATTTCCATCAGCGCTTATTTGAGCTGTCCTATACGAT 4212
1543 .....ThrTyrAspThrValAsp 1548

```



```

4213 GCGCGTCGCGCAAAAGTGGCAAGCGCGGATACCGCGCGATGTCGCA 4262
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1549 PheValSerGlyAspIysAspThr...ThrSerValThrValGluSerIy 1564
4263 GGAATTCGCGCAAAACCGCGAGTGGGAATGGCGGTAAACGCGCAATC. 4311
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1564 sAspAsnGlyLysAspThrGluThrGluValIleGlyAlaIleThrSerValI 1581
4312 .....AAAGTTTCACGCTGTCGCCATCCACGCTGCCGCC 4344
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1581 IeLysAspHisAsnGlyLysLeuPheThr..... 1590
4345 GCCAAGCGCGCGCAATGGAAAGCGCGACACAGCGCGGCATC 4386
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1591 .....GlyLysGluLeuLysAspAlaAsnAsnGlyVal 1602
seq_name: /cgn2_6/prodata/1/laa/5A.COMB.pep:US-08-409-995-4
seq_documentation_block:
  Sequence 4, Application US/08409995
  Patent No. 5646259
  GENERAL INFORMATION:
  APPLICANT: Barenkamp, Stephen I.
  APPLICANT: St. Gene III, Joseph W.
  TITLE OF INVENTION: Haemophilus Adhesion Proteins
  NUMBER OF SEQUENCES: 6
  CORRESPONDENCE ADDRESS:
  ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
  STREET: Four Embarcadero Center, Suite 3400
  CITY: San Francisco
  STATE: CA
  COUNTRY: USA
  ZIP: 94111-4187
  COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/409,995
  FILING DATE: 24-MAR-1995
  CLASSIFICATION:
  ATTORNEY/AGENT INFORMATION:
  NAME: Silva, Robin M.
  REGISTRATION NUMBER: 38,304
  REFERENCE/DOCKET NUMBER: A-61053/RFT
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 781-1989
  TELEFAX: (415) 398-3249
  TELEX: 910 277299
  INFORMATION FOR SEQ ID NO: 4:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 1912 amino acids
  TYPE: amino acid
  STRANDEDNESS: double
  TOPOLOGY: unknown
  US-08-409-995-4
alignment_scores:
  Quality: 248.50 Length: 1577
  Ratio: 0.350 Gaps: 70
  Percent Similarity: 44.959 Percent Identity: 19.784
alignment_block:
  US-09-303-518D-653 x US-08-409-995-4 ..
Align seq 1/1 to: US-08-409-995-4 from: 1 to: 1912
7 ACAACGCAAAAGCGACACGAA.....ACACACGCAAAAGCCC 47
  ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
245 ThrIyspLysAsnThrLeuAspValValLeuThrAlaIysGluAsn** 261

```

```

837 AAGTGGTTAAATGCGGTATTCGCAACGACGACCCCTATATA...G 883
      ::::: ::::: ::::: ::::: :::::
526 .....ValThrLysAspGlyIleIysAlaIleGlyAsnLysAlaIleIleHra 540
      ::::: ::::: ::::: ::::: :::::
884 GAAAAAGCAATGGCTTCGACGCTAGTTCGTAAGAATGGCTCTCATGATGAA 933
      ::::: ::::: ::::: ::::: :::::
540 snValAlaSerIleuAlaIleGlyAla.....TyrAspAsp 550
      ::::: ::::: ::::: ::::: :::::
934 ATCTTGGCTGGAGATACCATTCAGTATTTACGACACCATCAAAATGG 983
551 Ala..... 551
984 GAATACCTTTTAAACGCAATTAATATGCGGACGAAATATGATGCCA 1033
      ::::: ::::: ::::: ::::: :::::
552 .....AsnIleAspValLeuAsnAsnSerAlaThrAspLeu..... 563
1034 AACATTAACACTATTCCTACCTATAGATTTAAAAACGACACCGTTCAA 1083
      ::::: ::::: ::::: ::::: :::::
564 .....AsnArgHisValGluAspAlaIleLys.....Gly 573
1084 TTGTTTAATGTTTCTTTATCCGACGACGACGAAACCTGTTAT..... 1128
      ::::: ::::: ::::: ::::: :::::
574 LeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGluProLeuValThrAs 590
1129 .CATGTCGACGAGTGGGTCAACAGTTATTCACCCACGACTGAATATGAG 1177
      ::::: ::::: ::::: ::::: :::::
590 pSerThrAlaIleAlaThrValGluLysPleuArgLys..... 601
1178 AAAATATTTCTTATTTGACAAAGGAAAGGTAATGATTTACTTACCAGC 1227
      ::::: ::::: ::::: ::::: :::::
602 .....LeuGlyIlePheValSerThrLys 609
1228 AACATCAACGACGCGCGCGCTTTGTAATTTGAGGTAATTTTACGCT 1277
      ::::: ::::: ::::: ::::: :::::
610 Asn.....GlyThrLysGluGluSerAsn...GlnVal 619
1278 CTGCGCTTAAACACGAAACGTCGCAAGCGCGCGCTTATTCATCAGG 1327
      ::::: ::::: ::::: ::::: :::::
619 LysGlnAlaAspLysValLeuPheThrIleAlaGly..... 631
1328 ATGGCAGTACCGTTACTTGGAAAGTAACGGCGTGGCAACGACCGCTG 1377
      ::::: ::::: ::::: ::::: :::::
632 ..AlaIleAlaThrValThrSerLys..... 638
1378 TCCAAATGCGCAAGGACGCGCTGTTCAAGCCAAAGGGAACCA 1427
      ::::: ::::: ::::: ::::: :::::
639 SerGluAsnGlyLysHisThrIle..... 646
1428 AGGCTGCGTACGCGTGGCGAGGTAAGTCATTTAGATCAGCAGCGG 1477
      ::::: ::::: ::::: ::::: :::::
647 ....ThrValSerValAlaGluThrLys..... 654
1478 ACGATCAAGGCAAAAAACAGCCTTAGTAAGATCGCTGTGACGCGC 1527
      ::::: ::::: ::::: ::::: :::::
655 .....AlaAspCysGlyLeuGluLysAsp 662
1528 AGGGGACGCGTCACATGATCGCATATCATGTCACCCGCAAACT 1577
      ::::: ::::: ::::: ::::: :::::
663 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnValLe 679
1578 CTAATTTGGCTTTGCGGCGGACGTTTGATTTGAACGCGCATTCGCTT 1627
      ::::: ::::: ::::: ::::: :::::
679 uThrValGlyAsnAsnGlyThrAlaValThrLysGlyGlyPheGluThrV 696
1628 CGTCCACCGCATTCAAATACCGATGACGAGGCGGATGATGTCAC... 1674
      ::::: ::::: ::::: ::::: :::::
696 aLysThrGlyAlaThrAspAlaAspArgLysValThrValLysAsp 712
1675 .....CACAAATCAAGCAAGAA..... 1692
713 AlaThrAlaAsnAspAlaAspLysValAlaIleAlaThrValLysAspValAl 729
1693 .....TTCACCGTTTACCATTTACAGGCAATTAAGATTTACTTA 1729

```

```

729 aThrAlaIleAsnSerAlaIleAlaThrPheValLysThrGluAsnLeuThrT 746
      ::::: ::::: ::::: ::::: :::::
1730 CAACC...GGCAATTAACACACTTGATAGCAAAAAAGCAATTCCTAC 1776
      ::::: ::::: ::::: ::::: :::::
746 hrSerIleAspLysAspAsnProThrAspAsnGlyLysAspAlaLeu 762
1777 AAGGTTGG.....TTGGGACAGACATGACCA 1808
      ::::: ::::: ::::: ::::: :::::
763 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysVally 779
1809 AACGACGGCGCG.....CTCAATTCGATTTACC 1837
      ::::: ::::: ::::: ::::: :::::
779 SARAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuGluVal 796
1838 AACCGGAAGACGATGCGACTTCTACTGCTTCGCGGACCAAAATTA 1887
      ::::: ::::: ::::: ::::: :::::
796 ySThrAlaLysValSerAspThrLeuThrIleGlyLysAsnThrProThr 812
1888 AACGCG.....AATATCCGCAAAACGACCA 1916
      ::::: ::::: ::::: ::::: :::::
813 GlyLysIleThrAlaIleAlaThrProLysValAsnIleThrSerThrAla 829
1917 ACTGTTTTCAGCGCGACGACCGACCGACGCGCTCATCATTTAGAA 1966
      ::::: ::::: ::::: ::::: :::::
829 yLeuAsnPheAla...LysGluThrAlaAspAla.....SerGly 842
1967 GCGGCTGTCAAAAATGAAGTATCCACAAAGGAAATCGTGGGAC 2016
      ::::: ::::: ::::: ::::: :::::
842 eLysAsnValTyrLeuLysGlyIleAla..... 851
2017 AAGCATTTGATGACCGCACATTTAAAGCGGAAACTTCATATTCAG... 2064
      ::::: ::::: ::::: ::::: :::::
852 ThrThrLeuThrGluProSerAlaGlyAlaLysSerSerHisValAspLe 868
2065 .GGCGGACAAAGCGGTGTTCCCGCAATGTTGCCAAATGGA..... 2106
      ::::: ::::: ::::: ::::: :::::
868 uAsnValAspAlaIleThrLysSerAsnAlaIleSerIleGluAspVal 885
2107 ....GGCATTTGGCATTTAAGCAATCACGCCCAAGCAATTTGGGTGTC 2151
      ::::: ::::: ::::: ::::: :::::
885 euArgAlaGlyTyrPasnIleGlnGlyAsnGlyAsnAsnValAspTyrVal 901
2152 GCACCGCATCAAGCCACCAATCTGTACAGCTGCGACGCGGCGCT 2201
      ::::: ::::: ::::: ::::: :::::
902 AlaThrTyrAsp.....ThrValAsnPheThrAspSerThrGlyTh 916
2202 GACAACTTACCGCAAAAAACCATTACCGAGCATTA.....GTGATTG 2245
      ::::: ::::: ::::: ::::: :::::
916 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspVal 933
2246 CTTCAATTTGACAAAGCAGCATC..... 2268
      ::::: ::::: ::::: ::::: :::::
933 yIleGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 949
2269 .....AGAGGCAATGTCACGCTTGCGCA 2291
      ::::: ::::: ::::: ::::: :::::
950 ThrGlyLysAspLeuLysAspAlaAsnAsnGlyAlaThrValSerGluAs 966
2292 TCACGCTCATTTAATCTCACAGACTTGCACACTCAGCAAGGCAATTC 2341
      ::::: ::::: ::::: ::::: :::::
966 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIle 983
2342 GTGCGGCGGAGACAGCATATACGTTACG...CGCAACGCCACCCCA 2388
      ::::: ::::: ::::: ::::: :::::
983 spAlaValAsnLysSerGlyTyrArgValThrGlyGlnGlyAlaThrAla 999
2389 AACGCAACTCAGCTCGTG.....GGCAATGCCCAACGACATTTTA 2432
      ::::: ::::: ::::: ::::: :::::
1000 GluThrGlyAlaThrAlaValAlaAsnAlaGlyAsnAlaGluThrValThr 1016
2433 TCAAGCCACATTA.....AACGGCAAC.....ACATCGG 2461
      ::::: ::::: ::::: ::::: :::::

```

```

1016 rGlyThrSerValAsnPhelyAsnGlyAsnAlaThrThrAlaThrValS 1033
2462 CTTCCGACAAATGCTTATTAATCTAACCAACACGCCGTACAAACGGC 2511
      ::::::::::::::::::::
1033 eRlyAspAsnGlyAsnIleAsnValLysTyrAspValAsnValGlyLysP 1049
2512 AGCTGACGCTTCCGACACGCTTAGCAACGCAACGCAATCCGCACT 2561
      ::::::::::::::::::::
1050 GlyLeuLysIleGlyAspAspLysIleValAlaAspThrThrThrLe 1066
2562 C.....AACGCAATGTCCTCA..... 2580
      ::::::::::::::::::::
1066 uThrValThrGlyGlyLysValSerValProAlaGlyAlaAsnSerValA 1083
2581 .....GCCGATAGCAGTATTCATTGGA 2607
      ::::::::::::::::::::
1083 snAsnAsnLysLysLeuValAsnAlaGlyGlyLeuAlaThrAlaLeuAsn 1099
2608 AACAGCCGCTTACCGGAAA..... 2628
      ::::::::::::::::::::
1100 AsnLeuSerTrpThrAlaLysAlaAspLysTyrAlaAspGlyLysGcl 1116
2629 .....ATCAGCGCGCGC..... 2640
      ::::::::::::::::::::
1116 uGlyGlyuThrAspGlnGluValLysAlaGlyLysAspLysValThrPhelysA 1133
2641 .....AAGGATACGGCATTACAC 2658
      ::::::::::::::::::::
1133 laGlyLysAsnLeuLysValLysGlnSerGlyLysAspPheThrTyrSer 1149
2659 TTTAAAGACAGCAATGACGCGCGCTCG.....GCCACGA 2696
      ::::::::::::::::::::
1150 LeuGlnAspThrLeuThrGlyLeuThrSerIleThrLeuGlyLysThrAl 1166
2697 ATTAGCAAT.....TTAACTTGACAAACGCCACA 2728
      ::::::::::::::::::::
1166 aaGlnGlyArgAsnAspThrGlyThrValIleAsnLysAspGlyLeuThrI 1183
2729 TTTCACTCAATTCGCCCTATCGACAGATGCCGACGCGCAACCGGC 2778
      ::::::::::::::::::::
1183 LeThrLeu.....AlaAsnGlyAlaAlaAlaGly 1192
2779 AGTGGCGACATGCGCGCGCGCTTCGCGCGCTTCCCTAATATCCGT 2828
      ::::::::::::::::::::
1193 ThrAspAla.....SerAsnGlyAsnThrIleSerVal 1203
2829 TACGCGCGCACTTCGCGCAATCCGCTTCAACACGCTGACGTAACG 2878
      ::::::::::::::::::::
1203 LThr.....LysAspGlyIleSerAlaGlyAsnLysGluIleThrAsnV 1218
2879 GCAAATTGAACGCTCAGGAAACATTCGCTTATGCGGAACCTTCGCG 2928
      ::::::::::::::::::::
1218 allYs.....SerAlaLeuLysThr 1224
2929 TACCGCAGCGCAATTTGAAGCTGCGGAAAGTTCCGAGGCACTTACAC 2978
      ::::::::::::::::::::
1225 TyrLysAspThrGlnAsnThrAlaAspGluThrGlnAspLysGluPheN 1241
2979 CTTGGCTGTCAACAATACCGGCAAC...GAACCGCTAAGTCTGACGAA 3025
      ::::::::::::::::::::
1241 saLaIleValLysAsnAlaAsnGluValGluPheValGlyLysAsnGlyA 1258
3026 TGACGGTGTGGAAGAAAGCAACACACGCGTGTCCGAAATCTTAAT 3075
      ::::::::::::::::::::
1258 laThrValSerAlaLysThrAspAsn..... 1266
3076 TTCACCTGCAAAACGAACACGCTGATGCCGCGCATGGCTTATCAGCT 3125
      ::::::::::::::::::::
1266 ..... 1266
3126 TATCCGCAAAAGACGCGAGATTCCGCTGCATATCCGCTCAAAAGAACAG 3175
      ::::::::::::::::::::
1267 .....AsnGlyLysHisThrValThrIleAspValAlaGluAlaL 1280

```

```

3176 AGGTTTCCGACAAATCTGCGCAAGCGGAGAAACAGAGCCGCTTGACG 3225
      ::::::::::::::::::::
1280 ysaValGlyAspGlyLeuGluLys.....AspThrAspGlyLysIleLys 1294
3226 GCAAA.....CAGCACA 3239
      ::::::::::::::::::::
1295 LeuLysValAspAsnThrAspGlyAsnAsnLeuLeuThrValAspAlaTh 1311
3240 ACTTGGCCGCCAAACACACAGCGGAAAAAGCAACGCGCAAGCCTTGACG 3289
      ::::::::::::::::::::
1311 rLysGlyAlaSerValAlaLysGlyGluPheAsnAlaValThrThrAspa 1328
3290 CGCTGATTTGCGCGCGCGCAATGCCACCGCAAAAGCAGAAAGTGTGCC 3339
      ::::::::::::::::::::
1328 laThrThrAlaGlnGlyThrAsnAlaAsnGluThrGlyLysValAlaVal 1344
3340 GAACCGCGCGCGCAGCAGCGCGGGAATCCGCGCATTTGACAGCGCGCA 3389
      ::::::::::::::::::::
1345 Lys.....GlySerAsn...GlyAlaThrAlaThrGly 1354
      ::::::::::::::::::::
3390 GGAAGAGAAAGAAAGCGGTCAGCGGATAAAGACACCGCCTTGCGCAAC 3439
      ::::::::::::::::::::
1354 uThrAspLysLysValAlaThrValGlyAspValAlaLysAlaLea 1371
3440 AGCGCGAAGCG.....GAACCGCGCGCCTACACCGCC 3474
      ::::::::::::::::::::
1371 snAspAlaIleThrPheValLysValGluAsnAspSerAlaThrIle 1387
3475 TTCCCGCGCGCGCGCGCGCGCGCGGATTTGCGCGCAACCGCGCCCA 3524
      ::::::::::::::::::::
1388 .....AspAspSerProThrAspAspGlyAl 1396
3525 ACCGCAACCCCAACCGCAGCGCAGCTGATCAGCGCTTATGCAATAGCG 3574
      ::::::::::::::::::::
1396 aAsnAspAlaLeuLysAla***AspThrLeuThrLeuLysAlaGlyLysA 1413
3575 GTTTGAGTGAATTTTCCGCAACGCTCAACACGCTTTGCGCTACAGAC 3624
      ::::::::::::::::::::
1413 snLeuLysValLysArgAspGlyLysAsnIleThrPheAlaLeuAlaAsn 1429
3625 GAATTGGACCGGCTTTGCGCAAGCGCGCAACGCGCTTGGACAAG 3674
      ::::::::::::::::::::
1430 AspLeu.....SerValLysSerAl 1436
3675 CGCATCCGCGGACACCAACACTACCTTCGCAAGATTTCCGCGCTTACC 3724
      ::::::::::::::::::::
1436 aThrValSerAsp..... 1440
3725 GCCAACAAACGCACTGCGCCAAATCGTATGCAAGAAACCTGCGCAGC 3774
      ::::::::::::::::::::
1441 .....LysLeuSerLeuGlyThrAsnGlyAsnLysValAsnIleThrSer 1455
3775 GGGCGCGCTGCGCATCCTGTTTCG...CACACCGGACCGGA..... 3813
      ::::::::::::::::::::
1456 AspThrLysGlyLeuAsnPheAlaLysAspSerLysThrGlyAspAspAl 1472
3814 .....AACCTTTCGACGACGCGCATTCGGA 3838
      ::::::::::::::::::::
1472 aAsnIleHisLeuAsnGlyIleAlaSerThrLeuThrAspThrLeuAla 1489
3839 ACTCGCGCAGCGGCTTGGCCAGCGGTTCGCGGCAATFACGCAATCGCGC 3888
      ::::::::::::::::::::
1489 snSerGly.....AlaThrThrAsnLeuGlyLysGlyIleThrAs 1503
3889 AGGTTGACATCGCATCAGCGCGCGCGCGGCTTTAGTAGCGGACGCT 3938
      ::::::::::::::::::::
1503 paSngLysLys***AlaIleSerValLysAspValLeuAsnAlaGlyT 1520
3939 TTCAGACGCGCATCAGAGGCAAAATCCGCGCC..... 3970
      ::::::::::::::::::::
1520 rpaSnaValArgGlyValLysProAlaSerAlaAsnAsnGlnValGluAsn 1536

```



```
526 .....ValThrIysAspGlyIleLysAlaGlyAsnLysAlaLeuThr 540
540 .....ValThrIysAspGlyIleLysAlaGlyAsnLysAlaLeuThr 540
884 GAAAGCAATGCGCTCCAGCTACTTGGTAAAGATGGTCTATGATGAA 933
540 snValAlaSerGlyLeuAla.....TyrAspAsp 550
934 ATCTTGTGGAGATACCCATTCACTTCTTCAGAACCCATCAAAATGG 983
551 Ala..... 551
984 GAATACTTTTAAACGACATTAATAGCGCGAGAAAATCGATGCA 1033
552 .....AsnPheAspValLeuAsnAsnSerAlaThrAspLeu..... 563
1034 AACATAACACTATCTCTACCTTACCTTATAGATTAAAAACAGAACCGTTCAA 1083
564 ..AsnArgHisValGluAspAlaThrLys.....Gly 573
1084 TTGTTTAATGTTTCTTTATCCGAGACAGCAAGACCTGTTAT.... 1128
574 LeuLeuAsnLeuAsnGluLysAsnAlaAsnLysGlnProLeuValThrAs 590
1129 .CATGCTGAGGTGGGTCAACAGTTATGACCCAGACTGATATATGAG 1177
590 pSerThrAlaAlaThrValGlyAspLeuThrLys..... 601
1178 AAAATATTTCTTTATTGACAAAGAAAAGTGAATTGATCTATCCAGC 1227
602 .....LeuGlyTyrValSerThrLys..... 609
1228 AACATCAACCAAGCGCGCGGCTTGTATTGAGGTAATTACGCT 1277
610 Asn.....GlyThrLysGluGluSerAsn...GlnAla 619
1278 CTGCGCTAAACCAACGAAACGTCGCGCGGCTTCATATCATGTC 1327
619 LysGlnAlaAspGluValLeuPheThrGlyAlaGly..... 631
1328 ATGGCAGTACCGTTACTTGGAAAGTAACGGCGTGCACACCGCTTG 1377
632 ..AlaAlaThrValThrSerLys..... 638
1378 TCCAAAATCGCAAGCAGCGCTGCTGTTCAAGCCAAAGGGAACA 1427
639 SerGluAsnGlyLysHisThrIle..... 646
1428 AGCGTCGTCAGCGTGGCGCAGTAAGTCATCTTATGACGACGCG 1477
647 ....ThrValSerValAlaGluThrLys..... 654
1478 ACGATCAGGCAAAAACAGCCTTTAGTAATCGCGCTGTCACGCGC 1527
655 .....AlaAspGlySerGluLysAsp 662
1528 AGGGGAGCGGTCACATGATGCGATATCATGTTCAACCCGCAAACT 1577
663 GlyAspThrIleLysLeuLysValAspAsnGlnAsnThrAspAsnVal 679
1578 CTATTTGCGCTTTCGCGCGAGCTTGGATTGAGAGCGCATTCGCTT 1627
679 uThrValGlyLysAsnGlyThrAlaValThrLysGlyGlyPheGluThrV 696
1628 CGTTCACCGCATTCAAATACGATGAGGCGGATGATGTCAC... 1674
696 aLysThrGlyAlaThrAspAlaAspArgGlyLysValThrValLysAsp 712
1675 .....CACATCAAGACAAAGAA..... 1692
713 AlaThrAlaAsnAspAlaAspLysLysValAlaThrValLysAspValAl 729
1693 .....TTCACCGCTTACCATTTACGGAATTAAGATTTTCTA 1729
729 aThrAlaIleAsnSerAlaAlaThrPheValLysThrGluAsnLeuThr 746
1730 CAACC...GCCATTAACACCACTTGATAGCAAAAACAAATTCGCTAC 1776
746 hSerIleAspGluAspAsnProThrAspAsnGlyLysAspAlaLeu 762
1777 AACGGTTGG.....TTWGGCAGAAAGATGCAACCA 1808
763 LysAlaGlyAspThrLeuThrPheLysAlaGlyLysAsnLeuLysVal 779
1809 AACGAACGGCGG.....CTCATCTGATTAACC 1837
779 sArgAspGlyLysAsnIleThrPheAspLeuAlaLysAsnLeuVal 796
1838 AACCGGAAGAGCGGATTCGACTTCTGCTTCGCGCGCAACAAATTAA 1887
796 yThrAlaLysValSerAspThrLeuThrIleGlyLysAsnThrProThr 812
1888 AACGCG.....AAATTCAGCAACAAAGCGCA 1916
813 GlyLysThrThrAlaThrProLysValAsnIleThrSerThrAlaAsp 829
1917 ACTGTTTTCAGCGCGAGACCGACCGCCTCAATCATTTAGGA 1966
829 yLeuAsnPheAla...LysGluThrAlaAspAla.....SerGlyS 842
1967 GCGGGTGTCAAAATGAAAGTATCCACAGAGAAATGCTGTGGAC 2016
842 eLysAsnValTyrLeuLysGlyLeuAla..... 851
2017 AACGATTGATCGACCGCATTTAAAGCGAAACTTCATATTCG... 2064
852 ThrThrLeuThrGluProSerAlaGlyAlaLysSerSerHisValAsp 868
2065 .GCGGCAACAGCGGTGTTCCGCAATGTCGCAAGTGA..... 2106
868 uAsnValAspAlaThrLysLysSerAsnAlaIleSerIleLysVal 885
2107 .....GGCGATTGGCATTTAAGCAATCAGCCCAAGCATTTTCGGTGC 2151
885 euArgAlaGlyTyrAsnIleGlnGlyAsnGlyAsnValAspVal 901
2152 GCACCGCATCAAGCCACACATCTGTACAGTTGCGAGCGGCT 2201
902 AlaThrTyrAsp.....ThrValAsnPheThrAspSerThrGlyTh 916
2202 GACAGTTGTACCGAAACCAATACCGACGATTA.....GTGATG 2245
916 rThrThrValThrValThrGlnLysAlaAspGlyLysGlyAlaAspVal 933
2246 CTTCATTGACGCAAGCCGACATC..... 2268
933 yLeuGlyAlaLysThrSerValIleLysAspHisAsnGlyLysLeuPhe 949
2269 .....AGAGCAATGTCAGCTTCGCGA 2291
950 ThrGlyLysAspLeuLysAspAlaAsnGlyAlaThrValSerGluAs 966
2292 TCACGCTCATTTAAATCTCAGAGACTTGCACACTCAACGGCAATCTTA 2341
966 pAspGlyLysAspThrGlyThrGlyLeuValThrAlaLysThrValIle 983
2342 GTGACGCGGAGACACGCACTATACGTTAGC...CGCAAGCCGACCCCA 2388
983 sPalaValAsnLysSerGlyTyrPalaValThrGlyGluLysAlaThrAla 999
2389 AACGGAACCTCGCCTCGTG.....GGCAATGCCAAGCAACATTTAA 2432
1000 GluThrGlyAlaThrAlaValAsnAlaGlyAsnAlaGluThrValThrSe 1016
2433 TCAAGCCACATTA.....AACGCAAC.....ACATCGG 2461
1016 rGlyThrSerValAsnPheLysAsnGlyAsnAlaThrAlaThrValS 1033
```



```

1537 ILAspHeValAlaThrTyrAspThrValAspHeValSerLysPly 1553
3997 AGATACGGCGAGGTTTCGGCGATTCGGCATGCAACCGCATTCGGCGC 4046
1553 SASpThrSerValThrValGluSerLysAspAsnGlyLysArgHrg 1570
4047 AACGGCTATTTCGCCAAAAGCGGAT 4075
1570 LuValLysIleGlyAlaLysThrSerVal 1579

```

seq.name: /cgn2_6/prodata/1/1aa/6B_COMB.pep:US-09-413-814-78

```

seq_documentation_block:
; Sequence 78, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloeker, Helmut
; APPLICANT: Cino, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hoffe, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; EARLIER FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 78
; LENGTH: 882
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-78

```

alignment_scores:

Quality:	245.00	Length:	844
Ratio:	0.714	Gaps:	42
Percent Similarity:	40.640	Percent Identity:	24.526

alignment_block:

US-09-303-518D-653 x US-09-413-814-78 ..

```

Align seg 1/1 to: US-09-413-814-78 from: 1 to: 882
2539 GCAAGCTAAGCATTCG.....GCACTCAACGG 2567
75 AlMeIleIleIyHisSerLeuGlyGluLeuValAlaIalacysValaIagI 91
2568 CAATGCTCCCTAGCCGATTAAGCAGTATTCATTTCATTTGAAAACAGCCGCT 2617
91 yValPheThrLeuGluAspAlaMetSerLeuValIleAspArgGlyArgP 108
2618 TTACCGGAAAAATCAGCGCGGCAAGGNTACGGCATTTACACTTA..... 2661
108 heMetGlnAlaAlaProAlaGlySerMetLeuAlaValGlyLeuProAla 124
2662 AAAGACAGCAATGACGCTGCCGTCGGCGACGCAATTAAGCAATTTAA 2711
125 AlaAspValGluGlyLeuLeuGluAlaGlyLeuGluIleAlaIaValas 141
2712 CCTTGACACGCACTTACACTCAATTCCG..... 2743
141 nserProLysLeuThrValValaIagIy..ProAlaSerAlaIaIeArgasp 157

```

```

2744 .....CCTATCGACAC 2754
158 LeuAlaIaIaArgLeuGluAlaIaArgGluValaIaPheAlaIaArgProLeuGluThr 174
2755 GATGCGCGAG.....CGCGCAACCGCGCAG 2780
174 rSerHisAlaPheHisCysAlaLeuIleAspGlyAlaValaIaIaPhePhe 191
2781 TCGCGGAGATGGCGCGCGCGCGCGCGCGCG.....TCGCTATTAAT 2824
191 euGluSerValaIaArgAlaIaArgLeu..SerProProGluIleIleProVala 207
2825 C.....C 2826
207 lSerAsnValThrGlyAlaLeuLeuThrAspAlaIaIaThrAspProA 224
2827 GTTACGCGCGCAATTCGGCAGATCCGTTTCACACGCTGACGCTAA 2876
224 lAtyThrAlaIaIaArgHisLeuArgGlnProValaIaArgPheSerAspGlyVal 240
2877 CGGCAATTTGACGCTCAGGCAAC...ATTCCGCTTATGTCGCACT.. 2921
241 GluAlaLeuPheAlaSerGlyHisAlaLeuPheLeuGluValGlyProG 257
2922 .....CTTCGGCTACCGCGCGCAATTTGAAGCTGGCGGAAGTTC 2964
257 yArgGlyLeuThrThrLeuValaIaArgGlnThrLeuAlaGly..ArgGlyGly 273
2965 GAAGGCACTTACACCTTGGCTGCACAAATACCGGCAACGACCGGTAAG 3014
274 AlaAlaIaIaLeuSerLeuGlySerThrHisAlaIaSerIleProAlaSe 290
3015 TCTGCA.....GCAATTGACG 3030
290 rLeuAlaGluAlaLeuGlyGlnLeuThrGluAlaGlyHisAlaValaIaSpr 307
3031 GTAGTGAAGGAAAAAGACAACACACGCTGCGCAAAATGTTAATTTCAC 3080
307 rPThrAlaIaArgProArgGlyArgProProAlaIaArgSer..... 320
3081 CCTGCAAAACGACACGTCGA.....TGCGGCGCAT 3112
321 AlaCysArgArgThrArgSerSerGlyArgGlyThrLysSerArgAlaI 337
3113 GGCGTTATCAGCTTATCCGCAAGACGCGAGTTCGCTGCATTAATCCG 3162
337 AlaAlaIaProProProProSerArgArgArg...ProThrArgProSerA 353
3163 GTCAAGAACAAAGAGCTTCCGACAACCTCGCGCAAGCGGGAACAAACGA 3212
353 rGProArgArgThrProSerArgArgArgArgSerProAlaIaIaThrPro 369
3213 GGCGGCTTGAAGGCAAAACAGCAACACTGCGG..... 3247
370 Gly..ProArgSerProArgArgArgProSerProSerAlaIaArgSerPro 385
3248 .....CNAACACAGCGCGAAAAAGAC 3270
386 AlaIleThrPgluArgProProAlaIaArgArgArgArgArgArgArgArg 402
3271 AACGCGCAAGCTTGAAGCGCTGATTGCG.....CGGGGCGCAATGC 3314
402 gLeuLeuArgProAlaIaArgArgLeuAlaIaArgArgArgArgArgArgArg 419
3315 CACGCAAAAAGCAGAAAGTGTTCGCAACGCGCGCGCGGACGAGCGGG 3364
419 roAspProArg.....AlaAlaIaIaArgArgArgArgArgArgArgArg 432
3365 AAAATGCGCGCATTTACAGCGCGGAGAGAGAAAAAAGCGGTGCAGCG 3414
433 AlaLeuProGlyAlaAspGlyArgGlyAlaIaArgGlyProGlyArgArgArg 449

```

```

3415 GATTAAGACACCGCTTGGCGAAMACGCGGAAACCGCGCGCG 3464
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 g.....AlaProArgArgGlyAlaIaIa.....AlaArgV 459
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3465 TACGACCGCTTCCCGCGCGCGCGCGCGCGCGGATTGGCGAAC 3514
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
459 aLrProProAspArgAlaAspProArgArgProAlaProAlaVala 475
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3515 C...GCACGCCCCAACCGCAACCGCGAGCGAGCTGATCAGCGCT 3561
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
475 lArgAlaIaIaProAlaValaIaArgArgSerAlaIaArgIaIaPro 491
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3562 TATGCCATACGCGTTGAGTGAATTTCCGCGACGCTCAACGCGTTT 3611
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
492 .....GlyValGlnHisPrio 496
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3612 CGCGCTACGAGCAATTTGACG...CGTGTTCGCGAGAGCGCGCGCA 3658
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
497 ValArgGlyAlaArgAspGlyProAlaIaArgArgArgAlaProPro 513
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3659 ACGCCGTTTGACACGCGGATCGG..... 3683
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
513 nProValaIaArgAspArgAlaIaIaProArgGlyAlaIaIaAspAspValaGlna 530
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3684 .....GGACACCAACACTACGCTTCGCAAGATT..... 3713
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
530 rGlnGlnGlyArgArgAlaValProGlyArgArgAlaArgGlyProGlyAla 546
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3714 .....CGCGCGCTACG..... 3725
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
547 ValProAspValaIaArgProArgAlaIaProArgGlyArgGlyProArgArg 563
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3726 .....CCACAA 3732
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
563 yGlyIaIaGlyAlaIaArgProArgGlyIaIaProArgAlaValaIaArgProGln 580
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3733 ACCGACCTGGCCAAATCG.....TATGACAAACCTCGGAG... 3773
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
580 roArgAlaIaIaProProArgAlaGlyHisProGlnAlaIaProArgAlaIa 596
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3774 .....CGGCGCGCTCGCATCT.....GTTTTCGCACA 3802
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
597 ArgHisArgAlaIaIaArgAlaIaProArgAlaIaArgArgLeuValGlyArg 613
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3803 ACCGACCGGAAACCTTCGACGAGCGATCGCAACG.....GGCA 3846
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
613 gArgLeuArgArgArgAlaIaArgArgAlaIaIaArgArgLeuArgAlaGlyP 630
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3847 CGGCTTGCCGACGATG.....CGTTTCGGCAATACGCAAT 3884
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
630 roAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 646
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3885 CGGACGTTTCGACATCGCATACGCGCGCGCGGCTTTAGTACGCGCA 3934
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
647 SerProAlaGlyAlaIaValaIaIaArgArgGlyProGlyGly.....ArgAl 661
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3935 GCCTTTC..... 3941
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
661 aProValaIaIaAspArgAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 678
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3941 ..... 3941
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
678 lAgIaIaArgProArgAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 694
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3942 .....AGACGATCAGAGGAAATCGCGCGCGCTGTCATACG 3995
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
695 GluProArgArgArgGlyArgAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 711
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
3986 GCATTTCAGCAGATACGCGCGAG.....TTTCGGCGGATTCGGCATC 4029
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
711 gAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 728
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
4030 GAACGCGA.....CATCGCGC 4046

```

```

seq_name: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:US-08-169-927-2
seq_documentation_block:
; Sequence 2, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Mei
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/742,128
; FILING DATE: 08/09/91
; ATTORNEY/AGENT INFORMATION:

```


NAME: Spevack, A. David
 REGISTRATION NUMBER: 24,743
 REFERENCE/DOCKET NUMBER: 75,976
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 295-6759
 TELEFAX: (301) 295-1022
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1612 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-169-927-2

alignment_scores:
 Quality: 243.50 Length: 1471
 Ratio: 0.362 Gaps: 72
 Percent Similarity: 45.683 Percent Identity: 19.511

alignment_block:
 US-09-303-518D-653 x US-08-169-927-2 ..

Align seg 1/1 to: US-08-169-927-2 from: 1 to: 1612

```

634 AATAACCGGAAATTCATATCATATTCGAAGCGCATATCTTGGCTCGT 683
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
160 AsnSerArgSerSerSerSerSerSerSerSerSerSerSerSerSer 170
684 CGGTGGCATACCTTGGCAAAATGATGATGATGATGATGATGATGATGAT 733
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 .....AsnSerLysIleValAsnGly..GlyAsnGlyIleLeuAsni 184
734 TA.....GGTAGCGAAAAATTAAACATATGATGATGATGATGATGAT 777
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
184 LeThrAsnGlyPheIleGlnValSerAspAsnThrPheAlaGlyLeuLys 200
778 ACAGCAGGCTCATTTGGCGACAGCTGGCTCCACCAATGTTATCATGATGC 827
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 Thr...IleAsnIleAspAspCysGlnGlyLeuValLeuValSerLysAsn 216
828 CCAAAAGCAAAAGTGGTA.....A 847
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 AlaAlaIleAsnThrLeuAsnLeuGlnAlaGlyAsnThrIleAsnPr 233
848 TTAATGGGTATTCGAACAGCAACCCCTATATATGAGAAAAAGCAATGCG 897
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
233 heAsnGlyIleAspGlyThrGlyLysLeuValLeuValSerLysAsnGly 249
898 .....TTCCAGCTAGTCTCTAAAGATGATGATGATGATGATGATGAT 935
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
250 AlaAlaIleThrGlnPheAsnValThrGly..ThrLeuGlyGlyAsnLeuLys 266
936 CTTCGTGAGATATA...CCCATTCAGTATTCAGAACACATCAAAATG 982
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 LylleIleGlnLeuAsnThrAlaAlaValAlaGlyLysLeuIleSerLeu 282
983 GGA.....AATACCTTTTAAAGACATATATATGCGGAGGAGGAGAAA 1023
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
283 GlyGlyAlaAlaAsnAlaValIleGlyThr..AspAsnGlyAlaGlyAsn 298
1024 ATGCATCCCAACATTAACATATTCCTCTATATAGATTAAACACAG 1073
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
299 AlaAlaGly..... 301
1074 AACCGTCAATGTTTAAATGTTCTTTATCCGAGACAGCAGAGAACTG 1123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
302 .....PheIleValSerValAspAsnGlyAsnAlaAlaIleThr 314
1124 TTTATCATGCTGCAGTGGGTGCAACAGTTATGACACCAAGCAATGATAT 1173
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 IeSerGlyGlnValTyrAlaLysAsnMetValIleGlnSerAlaAsnAla 330

```

```

1174 GGAGAAATATTTCCCTTT.....ATTGACAAAGGAAAAAGT..... 1209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
331 GlyGlyGlnValThrPheGlnIleValAspValGlyLeuGlyGlyThr 347
1210 .....GAATTGATCTTACCACCAACATCA 1234
347 rThrAsnPheLysThrAlaSerSerLysValIleIleThrGlnLysSer 364
1235 ACCAAGCGCG.....GGCGTTTG..... 1254
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
364 snPheGlySerThrAsnPheGlyAsnLeuAspThrGlnIleValValPro 380
1255 .....TATTTGAGGCTAATTTACGCTCGCTTAAAAACACAGA 1295
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
381 AspThrLysIleLeuLysGlyAsnPheIleGlyAspValLysAsnGln 397
1296 A.....ACGTGGCAGCGCGCGGCGCTTCAATGATGAT 1326
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
397 yAsnThrAlaGlyValIleThrPheAsnAlaAsnGlyAlaLeuValSer 414
1327 .....GATGGCAGTACCTTACTTGAAGTAAACGCG..... 1359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
414 IeSerThrAspProAsnIleAlaValThrAsnIleAsnAlaIleGlnAla 430
1360 .....GTGGCAACGACCG 1373
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
431 GluGlyAlaGlyValValGlnLeuSerGlyIleHisIleAlaGlnLeu 447
1374 CCTGTCC.....AAATCGGCAAGGACGCGTGGTG 1405
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
447 glenGlyAsnGlyLysSerIlePheLysLeuAlaAspGlyThrValIle 464
1406 TTCACGCAAAAGGGAAC.....CAA 1428
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 snGlyProValAsnGlnAlaLeuMetAsnAsnAlaLeuAlaAla 480
1429 GGCCTGCTCAGCGTGGCGACGCTAAAGTCACTTCAATACGACGCGCA 1478
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
481 GlySerIleGlnLeu...AspGlySerAlaIleIleThrGlyAspIle 496
1479 CGATCAAGCGCAAAACACAGCTTATGAAATGCGTGGTGGCGCGCA 1528
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
496 yAsnGlyGlyValAlaAlaLeuGlnHisIleThrLeuAlaAsnAsp 513
1529 GGGGACGCGTG...CAACTGAATGCCGATATTCAGTCAACCGGAC... 1572
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
513 IeSerLysIleLeuAlaLeuAspGlyAlaAsnIleIleGlyAlaAsnVal 529
1573 .....AAACTGTATTTGCGCTTCGCGGCGGACGTTGGATTGACG 1616
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
530 GlyGlyAlaIleHisPheGlnAlaAsnGlyLysThrIle..... 542
1617 GCATTGCGTTTCGTCACCGCATTCAAATATTCAGTAAAGGCGCATGA 1666
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
543 .....LysLeuThrAsnThrGlnAsnIleVal 553
1667 TTGTCAACCAATCAAGACAAAGATCCACCGTTACCTTACGAGCAT 1716
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
553 al.....AsnPheAspLeuAspIleThrThrAspLysThrGlyVal 566
1717 AAAGTATTATTCACACCGGCAATACCAACAACTGATGATACAAAGAA 1766
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
567 ValAspAlaSerSerLeuThrAsnAsnGlnThrLeu..... 578
1767 AATGCTCAACAGGTGGTTGGCGAGAAAGATGCA...ACCAAAACGA 1813
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
579 .....ThrIleAsnGlySerIleGlyThrValValAlaAsnThrLysThr 594
1814 ACGGGCGGCTCAATCTGAATTCACCAACGGAAGAGCGGATCGCACTTA 1863
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
594 euAlaGlnLeuAsnIleGly.....SerSerLysThrIle 605
1864 CTCCTTCCGGCGGAGCAATTTAAACGCAATATACGCAAAACACG 1913

```



```

3387 GGAGGAGAGAGAAAAACGGGTGACGGGATTAAGACACCGCTTGGCGA 3436
      |||::: ::::: |||:::
1138 aglaspalaglavalasnalatthrthrthrthrthrthrthrleuysl 1155
      |||::: ::::: |||:::
3437 AACAGCGGAGAGCGGAAACCGG...CCGGTACACCGCTTCCCGCCG 3463
      |||::: ::::: |||:::
1155 leglnaspasnalasnalasnpheSerGlythrGlnAlatYthrleu 1171
      |||::: ::::: |||:::
3464 GCCCGCCCGCGCGGGATTTGGCGCAACCGACCCCAACCGCAAC 3533
      |||::: ::::: |||:::
1172 lIegInglYglYAlaArg..... 1177
      |||::: ::::: |||:::
3534 CCACCGCAGCGCGACCTGATCAGCGGTATGCCAATAGCGGTTGAC 3583
      |||::: ::::: |||:::
1178 .....PheAsnGlyThrleuGlyAlaPro 1186
      |||::: ::::: |||:::
3584 AATTTCGGCCGCTCAACAGCGTTTCGGCGTACAGGAGAAATTGGAC 3633
      |||::: ::::: |||:::
1186 snPheAlaValThrGlySerAsnIlePhe...ValLysTyrluLeu... 1200
      |||::: ::::: |||:::
3634 CGCGTGTTCGCGAAGACCGCGCAACCGCTTTGGACAAGCGGCATCG 3683
      |||::: ::::: |||:::
1201 .....IleArg 1202
      |||::: ::::: |||:::
3684 GGACACCAAAACACTACCGTTGCCAAGATTTCCGGCGCTACCGCAACA 3733
      |||::: ::::: |||:::
1202 gAspSer.....AsnGlnAspTyr..... 1208
      |||::: ::::: |||:::
3734 CGGACCTGGCCAAATCGGTATGCAGAAAACCTCGGACGGCGCGCTC 3783
      |||::: ::::: |||:::
1208 ..... 1208
      |||::: ::::: |||:::
3784 GGCATCTGTTTTCGCACACCGGACCGGAAACACCTTCGACGACGCA 3833
      |||::: ::::: |||:::
1209 ...ValleuThrArgThrAsnaspValleuAsnValThrThrAlaVal 1224
      |||::: ::::: |||:::
3834 CGGCAACTCGCGCAGCGCTTGCACCGCTCCGCTTCGGGCAATACGGA 3883
      |||::: ::::: |||:::
1224 lGlyAsnSerAlaIleAlaAsnAlaProIlyAl...SerGlnAsnIleSer 1240
      |||::: ::::: |||:::
3884 TCGGCA...GGTTCGACATCGGCATCGACCGCGGCGGCTTTAGTAC 3930
      |||::: ::::: |||:::
1241 ArgCysleuGlnSerThrAsnThrAlaAlaTyAsnAsnMetleuAl 1257
      |||::: ::::: |||:::
3931 G.....GCACCTTTCAGACGCGCATCAGAGCAAAAT 3962
      |||::: ::::: |||:::
1257 alysAspProSerAspValAlaThrPheValGlyAlaIleAlaThrAsp 1274
      |||::: ::::: |||:::
3963 CCGCGCGCGCGGTCTGCATTCAGGCAATTCAGCAAGATACCGCGCAT 4012
      |||::: ::::: |||:::
1274 hTrSerAlaAla.....ValThrThrValAsnleuAsnAspThrGlnLys 1288
      |||::: ::::: |||:::
4013 TCGCGGATTCGGCATCGAACCGCACATCGCGCAACGCGCTATTTCG 4062
      |||::: ::::: |||:::
1289 ThrGlnaspLeuSerAsnArg...leuGlyThrleuArgTyrlleSer 1304
      |||::: ::::: |||:::
4063 CAAAACGCGGATTCGATACGAAACGTCAATATGCGCACCGCGGCGCT 4112
      |||::: ::::: |||:::
1305 AsnAla.....GluThrSerAspAlaAla.....Glyse 1314
      |||::: ::::: |||:::
4113 TGCATTCACCGCTACCGCGCGGCGCATTAAGCATTAATTCATCAAC 4162
      |||::: ::::: |||:::
1314 rAlThrThrGlyAlaValSerSerGlyAspGlnAlaGlnValSerTyr... 1329
      |||::: ::::: |||:::
4163 CGCGGCAACACATTCATCAGCGCTTATTTGACGCTGCTCTATACGAT 4212
      |||::: ::::: |||:::
1330 .....GlyValThrPalalalysProPheTyrAsnIleAlaGlnGlnAsp 1343
      |||::: ::::: |||:::
4213 GCCCGTTCGGCG...AAAGTCGACACGCGC..... 4239
      |||::: ::::: |||:::
1344 LysLysGlyGlyIleAlaGlyTyrlValAlaLysThrThrGlyValVal 1360
      |||::: ::::: |||:::

```

```

4239 ..... 4239
1360 lGlyLeuAspThrLeuAlaSerAspAsnleuMetIleGlyAlaAlaIleG 1377
      |||::: ::::: |||:::
4240 .....GTCAATACCGCGCTATTCGGCGAGATTTCCGCAAAACCGCACT 4284
      |||::: ::::: |||:::
1377 lYleThrLysThrAspIleLysHISGlnAspTyrlLysLys..... 1390
      |||::: ::::: |||:::
4285 GCGGATGGGCGCTAAACCGCGCAATCAAGTTTCGACGCTGCGCTCA 4334
      |||::: ::::: |||:::
1391 .....GlyAspLysThrAspIleAsnGlyLeuSerPheSerleuTy 1404
      |||::: ::::: |||:::
4335 CGCTGCC 4341
      |||::: ::::: |||:::
1404 rGlySer 1406
      |||::: ::::: |||:::

seq_name: /cgn2_6/ptodata/1/laa/6B_COMB.pep:US-09-268-347-49

seq_documentation_block:
; Sequence 49, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-09-268-347-49

alignment_scores:
Quality: 237.00 Length: 1232
Ratio: 0.422 Gaps: 59
Percent Similarity: 45.536 Percent Identity: 19.805

alignment_block:
US-09-303-518D-653 x US-09-268-347-49 ..

Align seg 1/1 to: US-09-268-347-49 from: 1 to: 2314

340 GCACATAACGGCGCTATACCAATGTGATTTGGTCGCGAGGAGCA 389
      |||::: ::::: |||:::
25 SerHisSerGlyGlySerSerSerSerThrAlaGlyGlnValGlySerSe 41
      |||::: ::::: |||:::
390 TCCCGATCAGCACCGCTTTCTTACCAAATTTGTAAAGAAATATATATA 439
      |||::: ::::: |||:::
41 r.ProValIleArg..... 45
      |||::: ::::: |||:::
440 AAGCAGGAGCTAAGCGCATCTTATGGCGGATTAATATATATGCGCGT 489
      |||::: ::::: |||:::
46 .....LeuThrArgValAlaThrleuAlaIleleValIleGlyAl 59
      |||::: ::::: |||:::
490 TTGCACAATTTGTCA...CAGATGCAGAACCTGTTGAGATGACAGTTA 536
      |||::: ::::: |||:::
59 arThrleuAsnGlySerAlaTyrlaGlnAsnAsnSerLysIleAlaPheG 76
      |||::: ::::: |||:::
537 TATGATGGGTGAAT.....ACGCTGANT 562
      |||::: ::::: |||:::
76 lYThrThrGlyAsnAsnAspAsnAlaSerAlaSerAsnGlnAlaSerIle 92
      |||::: ::::: |||:::
563 TAAATTAATTAATTCGATTCGTTTCGATCGAGCAGCAGCAATATTCG 612
      |||::: ::::: |||:::
93 AlaIleGlySerleuAlaLysAlaHisAlaAsnGlnAlaIleAlaIle 109
      |||::: ::::: |||:::
613 CGGCTGATGAGACGACCAACCAAT.....AACCGGAAAGTTTC 650
      |||::: ::::: |||:::
109 yGly...SerLysProAspProArgAsnGlnAlaAlaAsnGlnLysAla 125
      |||::: ::::: |||:::

```

```

651 ATATCATTTGACAGCGCATATTCTTGCTGCTGCGATGACATACCTTTG 700
   : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
125 yserHisAlaLysGlyLysGlnSerIleAlaIleGlyLysPValLeuA 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
701 CACAAATGATCAGGTGCTGCGACATCACTTGTAGTGGCAAAAATTT 750
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
142 IAGLnu.....GlyAspAlaSerIleAlaIleGlySerAspSpleu 155
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
751 .....AACATAGCCCATATGCTTTTACCACAGAGGCTCATTT 791
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
156 TyrLeuAspArgAsnSerThrAsnSerIleTyrProAsnGly..... 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
792 TGGCAGACAGTGGCTCACCATGTTATCTATGATGCCAAAGCAAAAGT 841
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
169 ..... 169
842 GGTTAATTAATGGGTATTCGAA.....ACAGGCAACCCCTATATAGA 885
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
170 ..LeuLeuSerThrLeuIleGlnAsnHisThrValLeuArgGlnIleArg 185
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
886 AAACCATGGCTTCCACGCTAGTTCGTAAAGATTGTTCTATGATGAAAT 935
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 AspSerAsnGlySerGlnLysTyrArgArgThr..... 196
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
936 CTTCCTGAGATACCCAT.....TCAGTATCTACAGAACCAATC 976
   ||| : : : : : : : : : : : : : : : : : : : : : : : : :
197 ....AlaIleGlnGlyHisAlaSerThrAlaValGlyAlaMetAlaTyrA 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
977 AAAATGGCAATACTTTTTCACACAAATATATAGCGCGAGCAAAAATC 1026
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
212 IalysGlnHisPhe.....AlaAsnAlaPheGlyThrArg 223
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1027 GATGCCAAACATTAACACTATTCCTACCTTATATAGATTTAAACACGAA 1076
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
224 SerThrAlaGlnGlyAsnTyrSerGlnIleLeuAsnAsnAsnAsnAsn 240
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1077 CGTTCAATGTTTATGTTCTTATCCGACAGACAGCAAGCAACCTGTT 1126
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
240 aGluLysGlyTyrThrIleAlaIleGlySerAsnAlaGlnAlaIleAsn 257
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1127 ATCATGCTGCA.....GGT 1140
   || |||
257 TyrGlyAlaLeuAlaLeuGlyAlaAspThrArgValAspLeuAspTyrGly 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1141 GGGGTCAACAGTTATCGACCCACAGCG.....AATATAGCAGA 1178
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
274 IleAlaLeuGlyTyrGlySerGlnIleLeuAsnAsnAsnAsnAsnAsn 290
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1179 AATATATTCCTTATTTGACAAAGCAAGTGAATGATGATTACTTACGACA 1228
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
290 nAsnLysAlaTyrValProGlnGlyAsnGlySerAsnIleLysSerSer 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1229 ACATCAACCAAGCGCGCGCTTTGTATTTGAGGGTAAATTTTACGGTC 1278
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
307 ysaLysIaThr.....GlyAsnGlyLeuPheSerIleGlySerSerThrIle 321
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1279 TCGCCTAAATAACAACGAAAGCTGGCAAGCGCGCTTCATATCAGAGA 1328
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
322 LysArgLysIleIleLeuAsnVal.....GlyAlaGlyTyrGlnAspThrAs 336
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1329 TGGCAGTACCGCTTACTTGG.....AAGTAAACGGCGTGGCAACGAC 1372
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
336 pAlaValAsnValAlaGlnLeuLysAlaValGlnAsnLeuAlaLysArg 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1373 GCCTGCCAAATTCGGC.....AAAGCAGCGCTGCTTCAAGCCAAA 1416
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
353 IiIleThrPheLysGlyAspAsnGlyThrGlyValLysLysLeu 369
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1417 GGGGAAACCAAGCGCTCGTCAGCGTGGCGAAGTAAAGTCATCTTACA 1466
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : :
370 GlyGlu.....ThrLeuThrIleLysGlyLysGlnIleGlnAlaAs 383

```

```

1467 TCAGCAGCGGACGATCAAGCAAGCAAAACAGCCCTTATGTAATTCGCT 1516
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
383 pLysLeuThrAspAsn.....AsnAsnIleGly 393
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1517 TGGTACAGCGGACGGGAGCG.....GTGCACTG..... 1545
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
393 aIValThrAspAsnAsnThrIleLysValLysLeuAlaLysAsnLeu 409
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1546 .....AATGCCATATTCAGTTCACCCGACCAACT 1577
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
410 SerGlyLeuGlnThrValSerThrLysAsnLeuThrAlaSerGlnLysVa 426
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1578 CTATTCGCTTTCGCGCGACGCTTGGATTGTAAGCGCATTCGTTT 1627
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
426 IThrValGlySerGlyAsnAsnThrAlaGlnLeuGlnSerGlyLeuT 443
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1628 CGTTC.....CACGCGATTCAAATATCC 1650
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
443 hrPheThrProThrThrAsnAlaSerThrAspLysThrValTyrGlyThr 459
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1651 GATGAAGGCGCATGATTGTCACCAACATCAAGACAAAGAAATCCACGCT 1700
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
460 AspGlyLeuLysPheThrAspAsnSerAsnThrAlaLeuGlnAspThr 476
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1701 TACCATTTACA.....GGCAATAAAGATATATCTACAA 1732
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
476 ArgIleThrLysAspLysIleGlyPheSerAsnLysAlaGlyThrValA 493
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1733 CCGGCATATACCAACACTTGGATAGCAAAAAGAAATTCCTTACACGCT 1782
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
493 spGluAsnLysProTyrLeuAspLysAspLysLeuLysValGlyAsn... 508
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1783 TGGTTTGGCGAAGAAAGATGCACCAACCAAGACGGCGGCTCAATCTGA 1832
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
509 .....SerThrLeuAsnAsnGlyGlyLeuThrValAs 519
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1833 TTAACCAACGCGAAGACGCGATCGACTTACTGCTTCCGCGGAGAA 1882
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
519 n.....AsnThrIle.....GlyGlySerA 526
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1883 AT.....TTAAGCGCAAT 1896
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
526 snLysGlnIleGlnValGlyAlaAspGlyIleLysPheAlaAspValAsn 542
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1897 ATCAGCAACCAACACGCAACCTGTTTTCACGCGGACAGCGACGCA 1946
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
543 ValAsnValSerAsnAlaIleLysPheGlyThrThrArgIleThrGln 559
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1947 CGCTTCAATTCATTTAGCAAGCGGCTGTCAAAAATGAGATATCCAC 1996
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
559 uGlnIleGlyPheAlaAspAla.....A 567
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
1997 AAGGAGAATTCGTGGGACACGATTTGATCGACCGCATTTAAAGCG 2046
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
567 spGlyLysValAspLysLysSerProTyrLeuAspLysLys..... 580
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
2047 GAAACCTTCATATTCAGGCGGACAAAGCGGTGTTCCCAATGTTCG 2096
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
581 .....GlnLeuGlnValGlyLysIleThrIleThrLysAspSerG 594
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
2097 CAAAGTGGAGGCGATTGGCATTTAGCAATCAGCCCAAGCACTT... 2142
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
594 yLleAsnAlaGlyAspGlnLysIleSerAsnValLysAspAlaThrAsp 611
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
2142 ..... 2142
611 spThrAspAlaValThrTyrLysGlnLeuLysGlnValGlnIleAspAla 627
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
2143 .....TTGCGTGTGCAACCGCAT.....CAAAGCCA 2168
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
628 AspGlyAlaLeuGlnSerPheSerIleArgAspGlnLysGlnGlnLys 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
2169 CACATCTGTACAGCTTCGACTGACGCGGTCTGACAAAGTTGACCGAAA 2218

```

```

644 ethrlleaserleuTYrSer...AsnglyAsnThrProAsnThrheg 660
2219 AACCATTACC.....GACGATAAGCATGCTTCATTGAGCAG 2259
660 IuThrIleThrPheAlaGlyGluAsnGlyLe...SerIleSerAsn 674
2260 ACCGACATCAGAGCAATGTCAGCTTCGCCGATCAGCTCATTTAAATCT 2309
675 AsplleAlaLysGlyLysValLysVal..... 683
2310 CACAGCATTCGCACACTCAACGGC.....AATCTTAGTCAG 2347
684 ...GlyLleAspProIleAsnGlyLeuThrThrProLysLeuThrValG 699
2348 GCGGAGAC.....ACGCACTATACGTTACGCCGACAGCCAGC 2385
699 LysThrAspLysAspGlySerThrGlnLeuValIleGluGlnAlaLaser 715
2386 CAAACGGCAACCTCAGCCTGCGGCAATGCCAGACACATTTAAT.. 2433
716 GlyAsnGlyThrLysAsnIleLeuArgLysValSerProThrLeuProse 732
2434 .....CAAGCAATTAAACGCAACACATCGG 2461
732 rIleThrAsnAlaGlyGlyValArgThrThrGlnGlnGlnAsnThrIle 749
2462 CTTCGGAC.....AAT 2472
749 hSerAspLysAspLysSerLysAlaAlaSerIleGlyAspIleLeuAsn 765
2473 GCTTCATTATCTAAGCAACAGCCGTCACAAACGGCAGTCAGC... 2520
766 ThrGlyPheAsnLeuLysAsnSerAsnSerValGlyPheValSerThr 782
2521 .....CTTCCGACACAGCT...AAGCAACAGTAA 2548
782 rTyraThrValAspPheIleAspGlyAsnAlaThrThrAlaLysVal 799
2549 GCCATTCGCCACTCAACGGC.....AATGTCTCCCTA 2580
799 hTyraAspGluThrAsnGlnThrSerLysValThrTyraSypAlaAsnVal 815
2581 GCCGATAAGCAAGTATTCATTGTAACAGCCGTTACCGGAATAAT 2630
816 AspGluLysThrIleGluLeuThrGlyAspAsnGlyLysThrAsnLys 832
2631 C.....ACGGCGCGCAAGATA 2647
832 eGlyValLysThrThrIleuThrThrThrAsnAlaAsnGlyLysAla 849
2648 CGGCATTACCTTAAAGACAGCAATGGACGCTGCCGCGGCGCAGGAA 2697
849 hrasnpheSerThrThrAspAsnAspAlaLeuValAsnAlaLysAspIle 865
2698 TTAGGCAATTTAAAC..... 2712
866 AlaGluAsnLeuAsnThrIleuAlaLysGluIleHisThrThrLysGly 882
2713 .....CTTG 2716
882 rAlaAspThrAlaLeuGlnThrPheLysValLysAspGlyAlaThr 899
2717 ACAACGCCACCATTAACACTCAATCCGCCATGACACGATGGCGCGC 2766
899 spAspGluThrIleThrValGlyLys.....AspGly 909
2767 GCGCAACCGCAGTGGCGAGATGCGCGCGCGCGCTTCGCGCCGCTTC 2816
910 ThrIleAsnGlyLysThrVal..... 916
2817 CCTATTATCCGTTAGCGCGCACTGCGCAGAAATCCCGTTTCAACAGC 2866

```

```

917 .....AsnThrL 919
2867 TGACGTTAAACGGCAATGTAACGGTCAGGAAACATTCGCTTTATGTCG 2916
919 eubLysLeuLysGlyGlu...AsnGly..... 926
2917 GAACCTCTCGGCTACCGCAGCGGCAATTTGAAGCTGCGGAAAGTTCCGA 2966
927 .....LeuThrValAlaThrAsnLysAs 934
2967 AGCCTACTTACCTTGCGCTGTACACATATACCGCAACGACCGATAGTC 3016
934 pGlyThrValThrPheGlyLleAsnThrGlnSerGly..... 946
3017 TCGACATATTGACGTAATGGAAGAAAGACACACACCGCTGTCGGA 3066
947 .....LeuLysAlaGlyAspSerThrThrLeuAsnLys 957
3067 .....AATCTTAATTTACCTGCAACAAACGACGTCGATCCGG 3107
958 AspGlyLeuSerIleLysAsnProAlaSerAsnGlnIleGlnValG 974
3108 CGCA.....TGGCGTTATCAGCTTATCCGCAAA..... 3135
974 yAlaAspGlyValLysPheAlaLysValAspLysGlyAsnSerSerThr 991
3136 .....GACGGGAGTTCGCCCTGCAT..... 3156
991 LylleAspGlyThrSerArgIleThrLysAspGlnIleGlyPheThrGly 1007
3157 .....AATCGGTCACAAACACAGCTTCCGCAACT 3191
1008 AlaAsnGlySerLeuAspThrThrLysProHisLeuThrLysAspLys 1024
3192 CGGCAAGCGCGAGAAACAGAGCGCCGCTGACGCAAAACAGCA... 3237
1024 u...LysValGlyGluValGluIleThrAsnThrGlyLleAsnAlaGly 1040
3238 .....CAACTGCGCGCCAAACACAGCGGCAAAACACAGCGCAAGC 3282
1040 LylLysLysIleThrAsnIleGlnSerGly...AspIleThrGlnAsn 1054
3283 CTGACGCGCTGATTCGCGCGCGCGC.....AATGC 3314
1055 SerAsnAspAlaValThrGlyLysValThrValTyraSpleuLysThrGlu 1071
3315 CACGCAAGCGAGCAAGTGTGCGAAGCGCGCGCGAGCA 3357
1071 uGluSerLysIleAsnSerAlaAlaLysThrAlaGlnAsnSer 1085

seq_name: /cgn2_6/plodata/1/aa/5b_comb.pep:US-08-968-685A-10

seq_documentation_block:
: Sequence 10, Application US/08968685A
: Patent No. 6214981
: GENERAL INFORMATION:
: APPLICANT: TUCKER, KENNETH
: TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
: TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS LLP
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30

```

210 155CCGTCAGTATGAGACGAACTCAATAC..... 639

10 A.GCCAAGGGGAAACCAAGGCTGGTCAGCGTGCGG 1447

980 sThAspAlaLysGlyAspLysAlaIleAsnPhSerValAsnSerGlyA 997
1448 ACGGTAAGTCATCTTAGATCAGCGGAGCATCAAGCGAAACAA 1497
997 spAspLysAlaLeuIleAsn.....AlaLysAsp 1006
1498 GCCCTTACTGAATCGCTTGTCAGCGC.....AGGG 1532
1007 lIeAlaAspAsnLeuAsnThrLeuAlaGlyIleArgAsnThrLysG 1023
1533 GACG.....GTGCACTGA 1546
1023 YThrAlaSpThrAlaLeuGlnThrPheGlnValLysLysValLysG 1040
1547 ATGCGGATATCATGTTCAACCCGCAAACTCTATTGCTTCGCGGC 1556
1040 snGlyAspAspAsnAspAlaAspThrIleThrValGlyLysAspAla 1056
1597 GGACCTTGATTTGACGGGCGATTCGTTCCACCGCATTCANAA 1646
1057 LysThrAsnGlnValAsnThrLeuLysLeu..... 1066
1647 TACCGATGAAGGGCGATGATGTGCACCAATCAAGCAAGATCCA 1696
1067LysGlyLysAsnGlyLeuAspIleGlnThrAsnLysAspGlyT 1081
1697 CCGTATACCAT.....ACAGCAATAA 1719
1081 hrValThrPheGlyIleAsnThrGlnSerGlyLeuLysAlaGlyAsn... 1096
1720 GATATTACACACCGGCAATACAACTGATGACAAAGAAAT 1769
1097AsnThrThrLeuAsnAsnGlyLeu...SerIleLysAsnTh 1110
1770 TGCGTACAAAC.....GGTTGTTGGCGAGAAAGATGCA 1804
1110 rAlaGlyAsnGlnGlnIleGlnValGlyAlaAspGlyValLysPheAla 1127
1805 CCAAAACGAAGCGCGCTCATCTGAATTACCAACCGGAGAGAGCGGAT 1854
1127 ysValAsnAsnGlyValVal..... 1133
1855 CGCATTTACTGCTTCCGCGGACAAATTTAAAGCGCAATATCAGCA 1904
1134GlyAlaGlyIleAspGlyThrThrArg 1143
1905 AACAAACGGCAACTGTTTTCAGCGC.....AGAC 1936
1143 eThrArgAspGlnIleGlyPheAlaGlyThrAsnGlySerLeuAspLys 1160
1937 CGACACCGCAC...GCCATCAATCATTTAGAGCGGGTGTCAAAATG 1983
1160 eLysPheIleAsnSerLysAspGlyIleAsnAlaGlyGlyLysIle 1176
1984 GAAGGTATCCACAGAGAAATCTGTGGACACAGCATTTGATGACCG 2033
1177 ThrAsnIleGlnSerGlyIleAlaGlnAsnSer..... 1188
2034 CACATTTAAAGCGAAACTTCATATTCAGCGCGACAGCGGTGTT 2083
1189AsnAspAlaValThrGlyLysIleTyr... 1198
2084 CCGGCAATGTTGCCAAGTGAAGCGGATGGCATTTAACCATCAAGCC 2133
1199AspLeuLysThrGlnLeuGlnLysIleSerSerThrAla 1212
2134 CAAGCAGTT.....TTGCGTGTCCGACCGCATCAAG 2165
1213 LysThrAlaGlnAsnSerLeuHisGlnPheSerValAlaAspGlnGln 1229
2166 CCACACATCTGTACAGCTTCGCGACTGGAGGCTG...ACAAGTGTGA 2212
1229 yAsnAsnPheThrValSerAsnProTyrSerSerTyrAspThrSerLysT 1246

2213 CCGAAAAACCATTTACC.....GACGATAAAGTGAATTCCTTCATG 2253
1246 hrSerAspValIleThrPheAlaGlyGlnAsnGlyIleThrThrLysVal 1262
2254 AGCAAGACCGCATCAGA..... 2271
1263 AsnLysGlyValValAlaArgValGlyIleAspGlnThrLysGlyLeuThr 1279
2272GGCATGTCAGCTTG 2287
1279 rProLysLeuThrValGlyAsnAsnAsnGlyLysGlyIleValIleAsp 1296
2288 CCGATCAGCTCATTTAAATCTCACAGACTTGCCACACTCAAGCGCAAT 2337
1296 eArgAsnGlnGlnAsnThrIleThrGlyLeuSerAsnThrLeuAlaAsn 1312
2338 CTTAGTGCAGCGGAGACCGACCTATACGGTTACCGCGCACGCCCA 2387
1313 Val.....ThrAsnAspLysGlySerValArgThrThrGln 1324
2388 AACCGCAACCTC.....AGCTCG 2407
1324 uGlnGlyLysIleIleLysAspIleAspLysThrArgAlaIleSerIleV 1341
2408 TGCGCAATGCCCAAGCAACATTTAATCAAGCCATTTAAAGCAACACA 2457
1341 alaAspValLeuSerAlaGlyPheAsn.....LeuGlnGlyAsnGly 1354
2458 TCGGCTTCGACATGCTTCA.....TTTAATCTAAGCAACAA 2495
1355 GluAlaValAspPheValSerThrTyrAspThrValAsnPheAlaAspG 1371
2496 CGCGGTACAAACGCGAGCTGACGCTTCCGACACGCTAAGGCAAC. 2544
1371 yAsnAlaThrThrAlaLysValThrTyrAspAspThrSerLysThrSer 1388
2544 2544
1388 ysValValTyrAspValAsnValAspAspThrThrIleGlnValLysAsp 1404
2544 2544
1405 LysLysLeuGlyValLysThrThrThrLeuThrSerThrGlyThrGlyAl 1421
2544 2544
1421 aAsnLysPheAlaLeuSerAsnGlnAlaThrGlyAspAlaLeuValLysA 1438
2545GTAAAGCAT...TCCGACATCAACGCAATGTCCTTA 2580
1438 lAserAspIleValAlaHisLeuAsnThrLeuSerGlyAspIleGlnThr 1454
2581 GCC..... 2583
1455 AlalysGlyAlaIleSerGlnAlaAsnSerSerAlaGlyTyrValAspAlaAs 1471
2584GATAAGCGATATTC...CATTTGAAACAGCGCTTTACCGGAA 2626
1471 pGlyAsnLysValIleTyrAspSerThrAspAsnLysTyrGlnAla 1488
2627 AAATCAGCGCGGCAAGCATACGGCATTCACCTTAAAGACAGCGAATG 2676
1488 yAsnAspGlyThrValAspLysThrLysGlnValAlaLysAspLysLeu 1504
2677 ACGTCCCGCTCGGCGACGCAATTAGCAATTTAACTTGACAAAGCCCA 2726
1505 ValAlaGlnAlaGlnThrProAspGlyThrIleuAlaGlnImetAsnValLys 1521
2727 CATTAACATCAATTCGCTATGCACAGCATCGCGAGCGCGCA... 2772
1521 sSerValIleAsnLysGlnGlnIleValAsnAspAlaAsnLysLysGlnGly 1538

```
2773 .....ACGGCAGTGGCGGAGATGGCGCGC 2799
1538 leAsngluAspsnAlaPheValLySgLyLeuGluLysAlaIaIsertsp 1554
2800 GCGCGTGGCGCGCTTCCCTATATCCGT..... 2829
1555 AsnLyStrLysAsnAlaIaValThrValGlyAspLeuAsnAlaIaIaI 1571
2830 ....ACGGCG.....CCAACTTGGCGAGATGCC 2854
1571 agLnthrProLeuThrPheAlaGlyAspThrGlyThrThrAlaLysSL 1588
2855 GTTAAACACGCTGACGGTAAACGGCAATTGAACGGTCAG...GGAACA 2901
1588 euGlyGlyThrLeuThrIleLysGly.....GlyGlnThrAspThr 1601
2902 TTCCGCTTATGCGAGACTCTCGGCGTACGCGACGGC..... 2940
1602 AsnLyStrLysAspsnAsnIleGlyValAlaIaGlyThrAspGlyPh 1618
2941 .AAATTGAAGCTGGCGAAAGTTCGGAAGCATTACACCTTGGCTGTCA 2989
1618 eThrValLysLeuAlaLysAspLeuThrAsnLeuAsnSerVal..... 1632
2990 ACAATACCGGACGAAACCGTAACTCGAGCAATTGACGTAAGTg... 3036
1633 .AsnAlaGlyGlyThrLysIleAspAspLysGlyValSerPheValAsp 1648
3037 .....GAAGAAAAGACACACACCGCTGCCCAAAATCTTAATTT 3077
1649 SerSerGlyGlnAlaLysAlaAsnThrProVal..... 1659
3078 CACCTTCGCAAAACGACACGTCGATCGCGCGATGCGCTTATCAGTTA 3127
1660 ....LeuSerAlaAsnGlyLeuAspLeuGly.....LysValI 1672
3128 TCCCCAAGACGCGGAG...TTCCGCTGCATATCCGTCAAGAACA 3174
1672 leSerAsnValGlyLysGlyThrLysAspThrAspAlaIaAsnValGln 1688
3175 GAGCTTCCGAC.....AACTGGCAAGCGCGGAGCAAAAC 3209
1689 GlnLeuAsnGlnIValaIArgAsnLeuLeuGlyLeuGlyAsnAlaGlyAsnAs 1705
3210 AGAGCGCGCTTGAACGCAAAACGACACAACTTCCGCGCAACACAG 3259
1705 pAsnAla.....AspGlyAsnGlnValAsnIleAlaAspIleLysLysA 1720
3260 CGAATAAAGACACGCGCAAAAGCTTGACGCGCTGATGGCGC..... 3303
1720 spProAsnSerGlySerSerSerAsnArgThrValIleLysAlaGlyThr 1736
3304 .....GGCGCATGCCACGCAAAAGCGCAAAAGTTCG 3338
1737 ValLeuGlyGlyLysGlyAsnAsnAspThrGlnLysLeuAlaThr.... 1751
3339 CGAACGCGCGCGGACGCGGGAATGCGCGCATTTATGACGCGG 3388
1752 .....GlyGlyValGlnValGlyVal..... 1758
3389 AGGAAGAGAAAAACGGGTGACGCGGATTAAGACACGCGCTTGGCGAA 3438
1759 .....AspLysAsp..... 1761
3439 CAGCGGAGCGGAACCGCGCGGTACACACGCGCTTCCCGCGCGCG 3488
1762 .....G 1762
3489 CCGCGCGCGCGGATTTGCCCAACCGACGCGCCCAACCGCAACCCAA 3537
1762 yAsnAlaAsnGlyAspLeuSerAsnValThrValLysThrGlnLysAspG 1779
3538 ..CCGCGCGCGACCTGATCAGCGCTTATGCCAATAGCGGTTTGAAGTAA 3585

.....
1779 lySerLySAlaLeuLeuAlaThrTyAsnAlaIaGlyGlnThrAsn 1795
3586 TTTTCCCGCACAGCTCAACAGCTTTTCCGCGTACAGACGAATTTGACCG 3635
1796 Tyr...LeuThrAsnAsnProAlaGlnAlaIle.....AspAr 1807
3636 CGTGTTCGCGAA..... 3648
1807 gIleAsnGluGlnGlyIleArgPhePheHisValaAsnAspGlyAsnGln 1824
3649 .....GACCGCGCAACGCGTGTGGACAAAGCGCATCCGGAC 3687
1824 lndProValValGlnGlyArgAsnGlyIleAspSerSerAla.....Ser 1838
3688 ACCAACTACCTACCGTTCGCAAGATTTCCGCGCTTACCGGCAAAACGA 3737
1839 GlyLysHisSerValAlaIleGlyPheGlnAlaLysAlaAspGlyGlnAl 1855
3738 CCGCGCGCAATCGGTATGCAAGAAACCTCGGACGCGCGCGTGGCA 3787
1855 aAlaValAlaIleGlyArgGlnThrGlnAlaGlyAsnGlnSerIleAlaI 1872
3788 TCCTGTTTCGCAACACCGACCGGAAC...ACCTTGACGACGCGCATC 3834
1872 leGlyAspAsnAlaGlnAlaThrGlnLysPglAspGlnSerIleAlaIleGlyThr 1888
3835 GCGCACTCGCGACGCGCTTGCACGCGTGCCTTTTCGGG..... 3873
1889 GlyAsnValValAlaGlyLysHisSerGlyAlaIleGlyAspProSerTh 1905
3874 .....CAATACGCGATCGCGAG..... 3891
1905 rValLysAlaAspAsnSerTySerValGlyAsnAsnAsnGlnPheThra 1922
3892 .....TTCGACATCGGC..... 3903
1922 spAlaThrGlnThrAspValPheGlyValGlyAsnAsnIleThrValThr 1938
3904 .....ATCAGCGCGCGCGCGGTTTGTAGTACGCGGACGCTTTC 3941
1939 GluSerAsnSerValAlaLeuGly.....SerAsnSerAlaIleSe 1952
3942 AGACGCGATCAGAGCGCAAAATCCGCGCGCTGCTGATTAACGCGCATTC 3991
1952 rAlaGly.....ThriHisAlaGlyThrg 1960
3992 AGGCAAGATACCGCGCAGTTTCGGC..... 4017
1960 lndAlaLysLysSerAspGlyThrAlaGlyThrThrThrAlaGlyAla 1976
4018 .....GGATTGCGCATGCAACCGCACATCGCGCAACGCG 4052
1977 ThrGlyThrValLysGlyPheAlaGlyGlnThrAlaValAlaValSe 1993
4053 CTATTTCGTCGCAAAAGCGGATTTACGATACGAAGCAATCAATATC.... 4098
1993 rValGlyAlaSerGlyAlaGlnArgGlnGlnAsnValAlaAlaGly 2010
4098 ..... 4098
2010 lndAlaSerAlaThrSerThrAspAlaValAsnGlySerGlnLeuTyLys 2026
4099 GCGACCGCGCGCTTTCGA..... 4116
2027 AlaThrGlnGlyIleAlaAsnAlaThrAsnGlnLeuAspHisArgIleH 2043
4117 .....TTACCGCTACCGCGCGGCGCATTAAGCAGATTAATTCATTTCA 4159
2043 sGlnAsnGluAsnLysAlaAsnAlaGlyIleSerSer..... 2055
4160 AACGCGCGCAACATTTCCATCATCAGCGCTTATTTGAGCGCTGCTTATACC 4209
```


934 GluAsnLeuAsnThrLeuAlaLysGluIleHisThrThrLysGlyThrAl 950
1365 AAGACGACCGCTGTCCAAATATGCGCAAGCAGCGTGTGTTGACAGCA 1414
1366 AAspThrAlaLeuGln.....ThrPheThrValLysLysV 962
1415 AAGGGAAACCAAGGCTCGGTACGCGGCGGCAAGTAAAGTCTTA 1464
962 aAlaSpLysAsnAsnAlaAsp..... 969
1465 GATCAGACGCGGACGATCAGCAAGCAAAACAGCTTTAGTGAATCG 1514
970 AspAlaAsnAlaIleThrValGlyGlnLysAsnAlaAsnGlnValAs 986
1515 CTTGGTCAGCGGACGAGGCGGATGCACTGAATGCGCATATACAGTCA 1564
986 nThrLeuThrLeuLysGlyLysAsnGlyLeuAsnIleLysThrAspLysA 1003
1565 ACCCGACAAACTCTAT.....TTGCGCTTTGCGGCGCA 1599
1003 snglyThrValThrPheGlyIleAsnThrThrSerGlyLeuLysAlaGly 1019
1600 CGTTTGGATTGAACGGGCAATTCGCTTTCGTTTC..... 1632
1020 LysSerThrLeuAsnAspGlyGlyLeuSerIleLysAsnProThrGlySe 1036
1633 .CAGCGCATTTCAAAATACCGATGAGGCGGATGATGTCACACAAATC 1681
1036 rGlnGlnIleGlnValGlyAlaAspGlyValLysPheAlaLysValAsn 1053
1682 AAGCAAAAGATCCACCGTTACCATTCAGGCAATAAGATATTACTACA 1731
1053 snaGlyValValGlyAlaGlyIleAspGlyThrThrArgIleThrArg 1069
1732ACGCGCAATACACAACTTGATAGCAAAAGA 1766
1070 AspIleIleGlyPheThrGlyThrAsnGlySerLeuAspLysSerLysr 1086
1767 AATTGCTACACGCGTGTGGTGGCGGAAAGATGCAACCAAAACGACG 1816
1086 oHisLeu.....SerLysAspGlyIleAsnAlaGly 1097
1817 GCGCGCTCAATCTGAATTACCAACCGGAAGAACGCGATCGCACTTACG 1866
1097 LysLysLysIleThrAsnIleGln..... 1104
1867 CTTTCGCGGAAACAATTTAAACGGCAATATCAGCAAAACAGCGCA 1916
1105 ...SerGlyIleIleGlnAlaAsnSerHisAspAlaValThrGlyGly 1120
1917 ACTGTTT.....T 1924
1120 sIleThrAspLeuLysThrGluLeuLysLysIleSerSerThrAlaL 1137
1925 TCAGCGGACAGCAGCAGCGGCTTACATCATTTAGAGACGGGCGG 1974
1137 ystrhrAlaGlnAsnSerLeuHisGluPheSer.ValAlaAspGluGln 1153
1975 TCAAAAATGGAAGTATCCACAGAGAGAAATCGTGGGACAAAGATTG 2024
1153 yAsnAsnPheThrValSerAsnProLysSerSerLysAspThrSerLys 1170
2025 GATCGACCGC...ACATTTAAAGCGGAAACTTCATATTCAGGCGGA 2070
1170 hSerAspValIleThrPheAlaGlyGluAsn.....GlyIle 1182
2071 CAAGCGGTGTTCCGCAATGTGCAAAAGTGAAGCGGATTCGCAATT 2120
1183 ThrThrLysValAsnLysGlyValValArgValGlyIleAspGlnThr 1199
2121 A.....AGCAATCAGCCCAAGCAG 2140
1199 sGlyLeuThrThrProLysLeuThrValGlyAsnAsnAsnLysGly 1216
2141 TTTTGGTGTGCGACCGCATCAAGCCACACAAATCTGTACAGTTGCGAC 2190
1216 LeValIleAspSerGlnAsnGlyGlnAsnThrIle..... 1227
2191 TGGACGGGTCTGACAAAGTTGACGAAACCAATTCAGCAGTAAAGT 2240
1228 ...ThrGlyLeuSerAsn...ThrLeuAlaAsnValThrAsnAspLys.. 1241
2241 GATTCCTTCATTTAGACAGCAGCATCAGAGCAATGTC..... 2280
1242 ...GlySerValArgThrThrGlu...GlnGlyAsnIleIleLysAspG 1256
2281AGCCTTCCGAT.....CAGCGTCAATTGA 2304
1256 LuAspLysThrArgAlaAlaSerIleValAlaSpValLeuSerAlaGlyPhe 1272
2305 AATCTCAGACGACTTGCACACTCAGCGC..... 2334
1273 AsnLeuGlnGly.....AsnGlyGlnAlaValAlaAspPheValSe 1285
2335AATCTAGTCAGCGGAGACGACGCACTATATACG 2368
1285 rThrTyraSpThrValAsnPheAlaAspGlyAsnAlaThrThrAlaLysV 1302
2369 TTACGCGCAACGCCACG..... 2385
1302 aThrTyraSpThrSerLysThrSerLysValValTyraSpValAsn 1318
2386CAAAACGG 2393
1319 AspAspThrThrIleGluValLysAspLysLysLeuGlyValLysThr 1335
2394 CAACCTCAGCCTGTGGC.....AATGCCAAGCAACATTTAATC 2434
1335 rThrLeuThrSerThrGlyThrGlyAlaAsnLysPheAlaLeuSerAsnG 1352
2435 AAGCCATTTAAACGGCAACACATCGCTTCGACAAATCTTCATTAT 2484
1352 IlnAlaThrGlyAspAlaLeuValLysAlaSerAspIleValAlaHisSer 1368
2485 CTAGCAACAACGCGCTACAAAACGCGACGTCTGACGCTTCGCAACAGC 2534
1369 LeuAsn.....ThrLeuSerGlyAspI 1376
2535 TAAGCAAAACGTAGCATTCGCGACTCAGCGCAATGTCTCCAGCGG 2584
1376 eGlnThrAlaLysGlyAlaSerGlnAlaAsnSerAlaGlyTyraVal 1393
2585 AT.....AAGCAGATTCAT.....TTGAAACAGCGCGC 2616
1393 spAlaAspLysLysLysIleValIleTyraSpSerThrAspAsnLysr 1409
2617 TTTTACGGAATAATCAGCGCGGCAAGATACGCAATACCTTAAGAA 2666
1410 TyrGlnAlaLysAsnAspGlyThrValAspLysThrLysGluValAlaLys 1426
2667 CAGCGAATGAGCGCTGCGCTGCGGACGCAATTTAGCAATTTAAACCTG 2716
1426 sAspLysLeuValAlaGlnAlaGlnThrProAspGlyThrLeuAlaGln 1443
2717 ACAAGCGCACCATTTACATCATTCGCGCTATTCAGAGATGCGGCGG 2766
1443 eCAsnValLysSerValIleAsnLysGluGlnValAsnAspAlaAsnLys 1459
2767 GCGCAA.....ACGCGAGTGGCGGAGA 2789
1460 LysGlnGlyIleAsnGluAspAsnAlaPheValLysGlyLeuGlyLysAl 1476
2790 TGGCGCGCGCGCGTGTGCGCGCTTCCTATTATTCGCTT..... 2829
1476 aAlaSerAspAsnLysThrLysAsnAlaAlaValThrValGlyAspLeuA 1493

```

2830 .....ACGCG.....CCACTCG 2844
      |||||
1493 snAlaValAlaGlnThrProLeuThrPheAlaGlyAspThrGlyThrThr 1509
      |||||
2845 GCAGAAATCCGCTTCAACACGCTGACGGTAACGCAATTCAGCGTCA 2894
      |||||
1510 AlAlaGlyLeuGlyGlnThrLeuThrIleGlyGly.....GlyG 1523
      |||||
2895 G...GGAACATTCGCTTATTCGGAACCTTCGCTACCGACGCGC 2940
      |||||
1533 nThrAspThrAsnLysLeuThrAspAsnIleGlyValAlaIleGly 1540
      |||||
2941 .....AAATGAACTGGCGGAAAGTTCCGAAGGCACTTACAC 2979
      |||||
1540 hrAspGlyPheThrValLysLeuAlaLysAspLeuThrAsnLeuAsnSer 1556
      |||||
2980 TTGGCTGCACATACCGGCAACGCAACCCGTAGTCTCGACGCAATGAC 3029
      |||||
1557 Val.....AsnAlaGlyGlyThrLysIleAspAspLysGlyAlaSe 1570
      |||||
3030 GGTAGT.....GAAGAAAAGACACACACCGCTGTCGAAA 3067
      |||||
1570 rPheValAspSerSerGlyGlnAlaLysAlaAsnThrProVal..... 1584
      |||||
3068 ATCTTAATTTCCACCTGCAAAACGACGTCGATGCCGCGCATGGCGT 3117
      |||||
1585 .....LeuSerAlaAsnGlyLeuAspLeuGlyGly..... 1594
      |||||
3118 TATACGCTTATCCGCAAGACGCGAG...TTCCGCTGCAATATCCGCT 3164
      |||||
1595 ...LysValIleSerAsnValGlyLysGlyThrLysAspThrAspAlaAl 1610
      |||||
3165 CAAGAACAGAGCTTCCGAC.....AACTCGGCAAG 3199
      |||||
1610 asnValGlnGlnLeuAsnGlnValArgAsnLeuLeuGlyLeuGlyAsn 1627
      |||||
3200 CGGGAACACAGAGCGCCCTGACGCAAAACAGCAACTGCCCCC 3249
      |||||
1627 laGlyAsnAspAsnAla.....AspGlyAsnGlnValAsnIleAlaAsp 1641
      |||||
3250 AACACAGCGCGGAAAAAGACACGCGCAACCTTGACGCGCTGATTGC 3299
      |||||
1642 IleLysLysAspProAsnSerGlySerSerAsnArgThrValIleLys 1658
      |||||
3300 GGCC.....GGCGCAATGCCACCGAAAGGCGAG 3328
      |||||
1658 saIaGlyThrValIleuGlyGlyLysGlyAsnAsnAspThrGlnLysLeu 1675
      |||||
3329 AAAGTGTGCGCAACCGCGCGGACGAGCGGGAATCCGCGCAT 3378
      |||||
1675 laThr.....GlyGlyIleGlnValGlyVal 1683
      |||||
3379 ATGACGCGGAGAGAGAAAAACGGGTGACGCGGATAAAGACCGC 3428
      |||||
1684 .....AspLysAsp..... 1686
      |||||
3429 CTTGCGCAACACGCGCAACCGGAGCCGCTACACCGCTTCC 3478
      |||||
1686 ..... 1686
      |||||
3479 CCCGCGCGCGCGCGCGCGGATTTCCGCAACCGACCGCCACCG 3528
      |||||
1687 .....GlyAsnAlaAsnGlyAspLeuSerAsnValTrpValLysThr 1700
      |||||
3529 CAACCCCAA...CCGACGCGGACCTGATCAGCCGTTATGCCAATAGCG 3575
      |||||
1701 GlnLysAspGlySerLysLysAlaLeuLeuAlaThrLysAlaIle 1717
      |||||
3576 TTGGTGAATTTTCCGCAACGCTGATTCGCGGTTCGACAGAGC 3625
      |||||
1717 yGlnThrAsnThr...LeuThrAsnAsnProAlaIleAlaIle..... 1730
      |||||
3626 AATTGGACCGCGTGTTCGCGAA..... 3648
      |||||

```

```

1731 .....AspArgIleAsnGlnGlnGlyIleArgPhePheHisValAsnAsp 1745
      |||||
3649 .....GACCGCGCAACCGCGCTTGGACAAACGG 3677
      |||||
1746 GlyAsnGlnGlnProValValGlnGlyArgAsnGlyIleAspSerSerAl 1762
      |||||
3678 CATCCGGGACACCAACACTACGCTTCCGAAGATTTCCGGCGCTACCGC 3727
      |||||
1762 a.....SerLysHisSerValAlaIleGlyPheGlnAlaLysAla 1777
      |||||
3728 AACAAACGACCTCGCCCAATCGGTATGCAAGAAAACCTCGGACGCG 3777
      |||||
1777 spGlyGlnAlaAlaValAlaIleGlyArgGlnThrGlnAlaGlnGln 1793
      |||||
3778 CGCGTGGCATCTGTTTTCGACACACCGACCGCA..... 3813
      |||||
1794 SerIleAlaIleGlyAspAsnAlaGlnAlaThrGlyAspGlnSerIleAl 1810
      |||||
3813 ..... 3813
      |||||
1810 alIleGlyArgThrAsnValAlaIleGlyLysHisSerGlyAlaIleGly 1827
      |||||
3814 .....AACACTTCGACGACGCGCATCGCAC 3840
      |||||
1827 spProSerThrValLysAlaAspAsnSerThr.....SerValGlyAsn 1841
      |||||
3841 TCGGCACGCGCTTCCGACGCGGCGTTCGCGCAATACGCAATCGGCGAG 3890
      |||||
1842 AsnAsnGlnPheThrAspAlaThrGlnThrAspAlaPheGlyValGlyAs 1858
      |||||
3891 GTTCGACATCGGCATCAGCGCGCGCGGTTCGATGACGCGCATTT 3940
      |||||
1858 n...AsnIleThrValThrGlnLysSerValAlaLeuGlySerAsn 1874
      |||||
3941 CAGACGCGCATCAGAGCAAAATCCCGCGCGCTGCTCATTCAGCAT 3990
      |||||
1874 erAlaIleSerAlaGly.....ThrHisAlaGlyThr 1884
      |||||
3991 CAGCAAGATACCGCGCGCGCTTCGCGC..... 4017
      |||||
1885 GlnAlaLysLysSerAspGlyThrAlaGlyThrThrThrAlaGlyAl 1901
      |||||
4018 .....GGATTCGCATCGACACCGCATCGCGGCAACGC 4051
      |||||
1901 aThrGlyThrValLysGlyPheAlaGlyIleThrAlaValAlaValS 1918
      |||||
4052 GCTATTTCGTCGCAAAAGCGGATTCGATACGAAACGTCATATTC... 4098
      |||||
1918 erValGlyLysSerGlyAlaGlnArgArgIleGlnAsnValAlaIleAlaGly 1934
      |||||
4098 ..... 4098
      |||||
1935 GluValSerAlaThrSerThrAspAlaValaAsnGlySerGlnLeuThrLys 1951
      |||||
4099 .....GCCACCGCGCGCGCTTCGATTCACACGCT 4126
      |||||
1951 saIaThrGlnSerIleAlaAsnAlaThrAsnGlnIleuAspHisArgIleH 1968
      |||||
4127 ACCGCGCGCGCATTAAGGACGATTAATTCATTAACCGCGGCAACACAT 4176
      |||||
1968 isGlnAsnGlnAsnLysAlaAsnAlaGlyIleSerSerAlaMetAlaMet 1984
      |||||
4177 TCCATCACGCGCTTATTTGACCTGTCTATACGATGCGCGCTTCGCGCA 4226
      |||||
1985 AlaserMetProGlnAlaThrIleProGlyArgSerMetValThrGly 2001
      |||||
4227 AGTCGGAAGCGCGTCAATACCGCGTATGCGGACGATTCGCGCAA 4275
      |||||
2001 yIleAlaThrHisAsnGlyGlnGlyAlaValAlaValGlyLeuSerLysL 2018
      |||||
4276 ..ACCGCGAGTGGCGAATG.....GGCTAAACGCGCAAAATC 4311
      |||||

```

```

2018 euSerSpasnglyIntPvalPheLysIleasnGlySerAlaAspThr 2034
4312 AAGGTTTCACGCTGCTCCAGCGTGGCGCCGCGCAAGGG 4353
2035 GlnGly.....HisValGlyAlaIaValGly 2043

seq_name: /cgn2_6/ptodata/1/lae/6B_COMB pep:US-09-413-814-79

seq_documentation_block:
: Sequence 79, Application US/09413814
: Patent No. 6225064
: GENERAL INFORMATION:
: APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
: APPLICANT: Bristol-Myers Squibb, Co.
: APPLICANT: Beyer, Stefan
: APPLICANT: Bloecker, Helmut
: APPLICANT: Cino, Paul M
: APPLICANT: Dougherty, Brian A
: APPLICANT: Goldberg, Steven L
: APPLICANT: Hottle, Gerhard
: APPLICANT: Mueller, Joachim
: APPLICANT: Reichenbach, Hans
: TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
: FILE REFERENCE: PCT/US 99/23535
: CURRENT APPLICATION NUMBER: US/09/413,814
: EARLIER FILING DATE: 1999-10-07
: NUMBER OF SEQ ID NOS: 107
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 79
: LENGTH: 1213
: TYPE: PRT
: ORGANISM: Strangium celluloseum
US-09-413-814-79

alignment_scores:
Quality: 211.00 Length: 693
Ratio: 0.751 Gaps: 33
Percent Similarity: 40.548 Percent Identity: 24.820

alignment_block:
US-09-303-518d-653 x US-09-413-814-79 ..
Align seg 1/1 to: US-09-413-814-79 from: 1 to: 1213

2562 CAACGGCAATGTCCTCCAGCGATAGAGCATTTTGAACA 2611
:::|||||: ||| |||: |||
5 ArgArgArgAlaAlaProHisLeuArgGlyAlaArgPro..... 17
2612 GCCGCTTACCGGAAATACCGCGCGCAAGATACGCAATTA 2661
|||:|||||: ||| |||: |||
18 .....AlaArgArgGlyAspArgArgGlyLeuAlaLeuArgSer 32
2662 AAGACAGCGAATGAGCGCTGCCGCGCGCAAGATTAAGCA 2711
|||:|||||: ||| |||: |||
32 spArgGlnArgAlaArgArgArg...GlyHisGly..... 42
2712 CTTGACAAAGCCACCATTAATTCGCTATCGACAGCATGCG 2761
43 .....GluGlyLeuGlyAlaGlyGly 49
2762 CAGGCGCGCAACCGGCAATGCGGAGATGCGCGCGCGCTT 2811
|||||:|||||: ||| |||: |||
49 YArgArgAlaArgArgAlaAlaArgArgArgArgLeuProAla 66
2812 GTTCCCATTAATCGTTACGCGCGCAACTGCGCAATTCGCT 2861
|| |||
66 roSerPro..... 68

2862 CACGCTGACGGTAACGGCAATTGAAACGGTCAACGTAATTC 2911
69 .....ProArgGlyAlaProGlnAlaPr 76
2912 TGTGGAACTCTTCGCTACCGGACGGCAATTTGAACTGG 2961
|||||: ||| ||| |||: |||
76 oProGlyAlaArgArgGlyProProArgProHisAlaValGly 91
2962 TCCGAAGCAGCTTACACCTTGCTCAACATACCGGCAAC 3011
|||||: ||| ||| |||: |||
92 ArgArgAspArgArgLeuAlaArgGlyHisAlaArgAlaArg 107
3012 AAGTTCAGCAATTTGACGGTAGTGAAGAAAGACACACCG 3061
|||||: ||| ||| |||: |||
108 ArgGlyArgAlaValAlaArgAlaAlaArgAla..... 119
3062 CCGAAATCTTAATTTCAACCTGCAAAAGACACGCTGCG 3111
|||||: ||| ||| |||: |||
120 .....ProAlaAlaArgArgGlyHisGlyArg..... 128
3112 TGCGGTATACGTTATCCGCAAGACGG.....CGAGTTCC 3149
|||||: ||| ||| |||: |||
129 .....ProGlyLeuArgArgLeuHisValArgValHis 139
3150 CTTGCATATCCGCTCAACAGACGCTTCCGCAAACTGCGCA 3199
|||||: ||| ||| |||: |||
140 GlyAlaAlaGlnGlyArgAspArgProProArgArgGly 156
3300 CGGAGAAACAGAGCGCGCTTGACGGCAAAACGACACACT 3249
|||: ||| ||| |||: |||
156 yProArgHisGlnProProLeu.....ArgA 165
3250 AAMCAGCGCGGCAAAAGACACGCGCA.....AAGCT 3284
|||: ||| ||| |||: |||
165 rArgArgProGlyGlyProGlyAlaArgAlaLeuValAla 181
3285 TGACGC.....GCTGATTCGCGCGCGCGCAATGCGCA 3322
|||||: ||| ||| |||: |||
182 ValGlyLeuArgArgValAlaArgAspAlaArgArgArg 198
3323 AGGCAAGAGTGTTCGCAACCGCGCGCGCGCGGCAAAAT 3372
|||||: ||| ||| |||: |||
198 pProArgProGlyProArgGlyGly...SerGlyAlaLeu 214
3373 GGCATTATGACGCGGAGACAGAAAGACGCTGACGGCAT 3422
|||: ||| ||| |||: |||
214 rGlyAlaArgAlaGlyAspArgValGlyLeuGlyProGly 230
3423 CACCGCTTGGCGAAACAGCGGAGCGGAAACCGCGGCT 3472
|||||: ||| ||| |||: |||
231 AspAlaHisGlyArgValProArgArgGlyArgPro... 246
3473 CCTTCCCGCGCGCGCGCGCG.....CGCGGAGATTCG 3513
|||||: ||| ||| |||: |||
246 lAlaProProArgHisAspGlnArgArgLeuAspProAla 263
3514 CCGGACCCCAACCGCAACCGCGGAGCGGCACTGATCAG 3563
|||||: ||| ||| |||: |||
263 rGSerHisProArgGlyLeuProArgAlaProArg..... 274
3564 TGCATATAGCGTTTGAGTAATTTCCCGCAGCTCAGAG 3613
|||||: ||| ||| |||: |||
275 .....ArgGlnProArgArgArgAspArg 282
3614 CCGTACAGAGCAATTTGACCGGCTGTTTCCGCAACCG 3663
|||: ||| ||| |||: |||
282 G.....GlyValAspLeuValAspArgProProAspArg 293
3664 GTTTCAGCAACCGGCACTCGGAGACCAACACTACGCT 3713
|||||: ||| ||| |||: |||
294 .....GlyArgArgProGly.....ValAlaGlnHis 302
3714 CCGCGCTACCGCGCAACGACGACTCG..... 3743

```

```
303 ProLeuArgProProAlaArgGluProAlaHisLeuArgAlaArgArgG1 319
3744 .....CCAAATCGGTATGCGAGAAAACCTC 3768
319 yAlaGlyAlaValSerAspArgGlyProArgArg.....AspProH 333
3769 GGCAGCGCGCGCTCGGCATCCTGTTTC..... 3797
333 ISArgArgAspArgArgAlaArgGluLeuAlaArgGlyAlaAsp 349
3798 .....GCACAAACCGACCGAACCCTTCAGCA.....CGGCA 3832
350 ProGlyAlaValProGluAlaProHisAspArgArgAlaValGlnAs 366
3833 TCGGCACTCGGACGCGCTGCCACGCGTCCGCTTTGCGCAATACGG 3881
366 pArgArgProGlyProLeuLeuArgArgAspHisArgAlaAlaArgA 383
3882 ..CATCGG.....CAGGTTGCATCGGCATCAGCGC 3911
383 IAlaHisArgProSerGlyGluAspProArgLeuProHisArgAlaArgArg 399
3912 GGGCGCGGGGTTTACTAGCGGACCTTTACAGACGCGCATCAGAGCAAA 3961
400 AspArgGlyArgProArgAlaAlaProLeuGlyArgAlaGlyGlyArgG1 416
3962 TCCGCGCGCGCGTCTGCATTTAGGCATTCAGGCAAGATACGCGCAGGT 4011
416 yGlyGluAspArgSerValGlyArgGluAla.....ProGlyArgV 430
4012 TTGGGCGGATTCGGCATCGACCGCACATCGCGCAGCGCGCTATTTCGT 4061
430 AlaArgArg...ArgArgArgArgArgGTPArgArgAlaProArgLeuArg 445
4062 CCAAAA.....AGCGATTACCGATACGAAAACGTCAATATCGCCACCC 4105
446 ProGluGluAlaAlaGlyValHisAspProArgGlyGlyArgArgProPr 462
4106 CGGGCGCTTCGATTCAACCGCTACCGCGCGCGCATTAAGCGAGATTATTC 4155
462 OGlyAlaAlaAlaGluArgGluArgGlnGlyProArgArg..... 476
4156 TTCAAACCGCGCACACATTCATCAGCCTTATTGAGCGTGCCTA 4205
477 ..ProAlaGlySerArgArgGlyArgAlaAla..... 486
4206 TACCGATCCCGCTTCGCGCAAAAGTCCGAACGCGCTCAATACCGCGGTAT 4255
487 ...ArgArgGlyArgArgProAlaAspArgAspArgAlaAlaHisArgLe 502
4256 TGGCGCAGGATTCGGCAAAACCCGCAATGCGGAATGGGGCGCTAAACGCC 4305
502 uGlyAlaAla.....ArgGlyGlyLeuAlaGlyArgGlyArgArgArgA 516
4306 GAATATCA.....AGTTTCACGCTGTCCCTCCAGCGTCCGC 4343
516 spArgGlnProLeuAlaArgAlaArgLeuHisLeuAlaAlaProArgAlaArg 532
4344 .....CG 4345
533 AlAlaAlaProAlaArgArgAlaHisArgGlyAlaArgAlaArgArgGlyArg 549
4346 CCAAGGGGCC..... 4355
549 gArgGlyAlaGlyArgValAlaAspArgProValProValProAspHisA 566
4356 ..GCAATTGGAAGCGCAGCACAGCGCGG 4382
566 rGAlaThrGlyAlaAlaAlaAlaArgArgGly 575
```

